Photographic Plates

Photographic plates were the workhorse of astronomy for almost 100 years. However, these plates were

- Low (~ 1%) efficiency detectors. (With hypersensitizing, this could be increased to almost 3%.)
- Non-linear detectors. (Some types more non-linear than others)

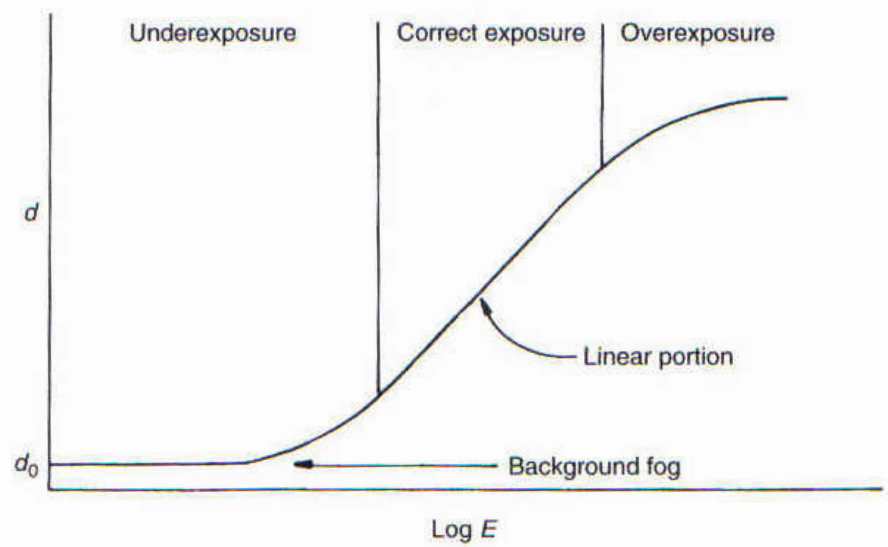


Fig. 3.2. The characteristic curve of a photographic emulsion. At very low values of the exposure E the density d tends to a constant low background value – a 'fog' which is a property of the plate rather than a measure of the exposure. Over a short range of exposure, the density varies linearly with exposure, but overexposure causes saturation and the density again ceases to give any information about the exposure.

Photographic Plates

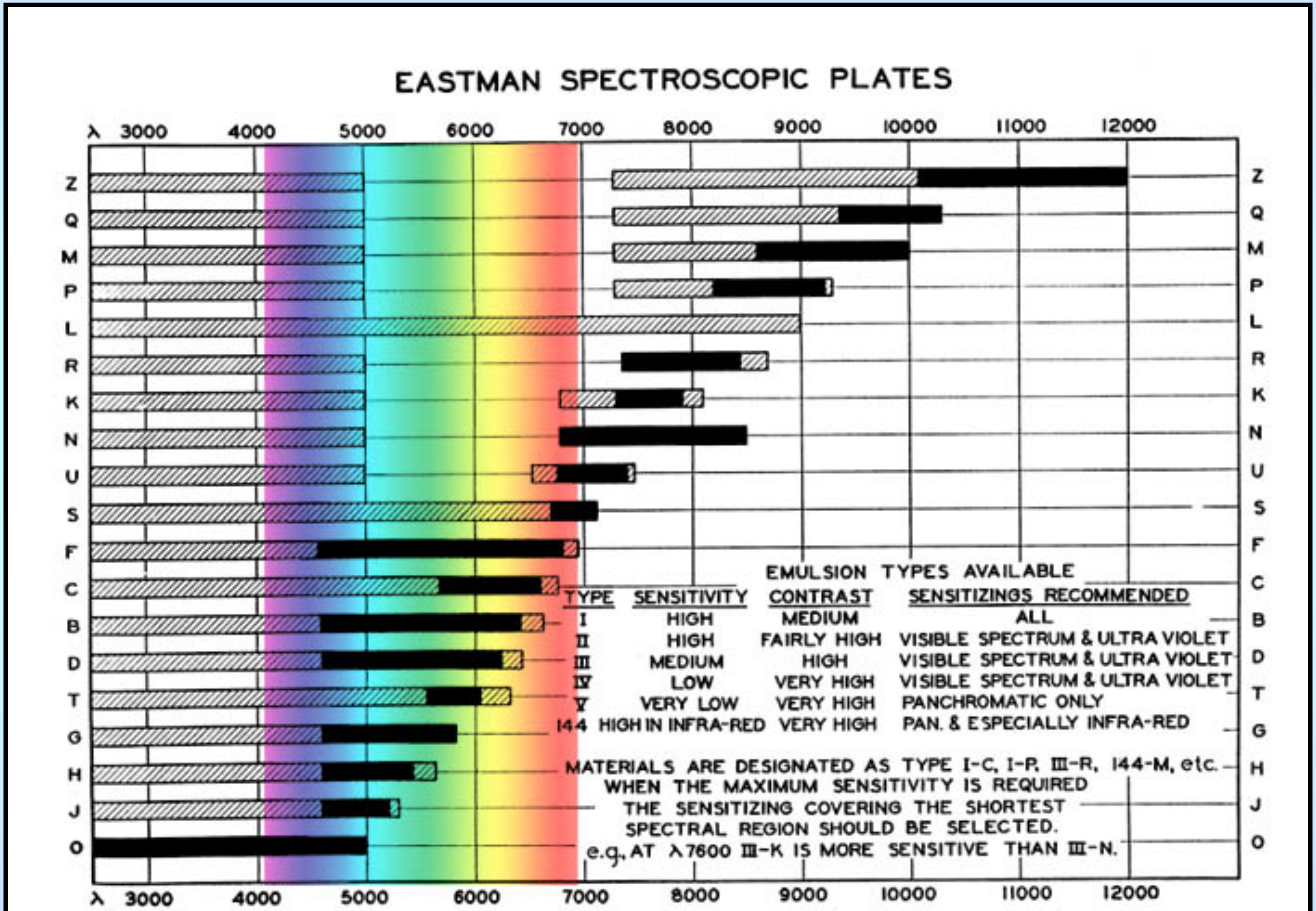
There were many types of photographic plates. Plates were defined via their a) linearity for astronomical observations, b) red response, and c) grain size. These were some of the most popular plate types:

Type	**Grain-size**	**Red limit**	**Name**
103a-O	coarse	5000 Å	B
IIa-O	medium	5000 Å	B
IIIa-J	fine	5500 Å	J
IIa-D	medium	6500 Å	V
103a-E	coarse	6700 Å	R
098	coarse	6900 Å	R
IIIa-F	fine	6900 Å	F
IV-N	fine	9000 Å	I

The letter defined the red-limit of the response, while the number or roman numeral described the grain size.

Photographic Plates

This was the list of plate types in 1937.



Cross-hatched areas show total sensitivity. Solid block areas show spectral regions for which type is especially valuable.

Photomultiplier Tubes

Prior to the mid-1980's, the highest efficiency detectors were Photo-Multiplier Tubes (PMTs). A photon dislodges an electron from a photo-cathode. The electron accelerates through an electric or magnetic field and strikes a dynode, which ejects more electrons, which fall to the next dynode, etc. A single incident electron can be amplified millions of times! The result is a pulse of electricity.

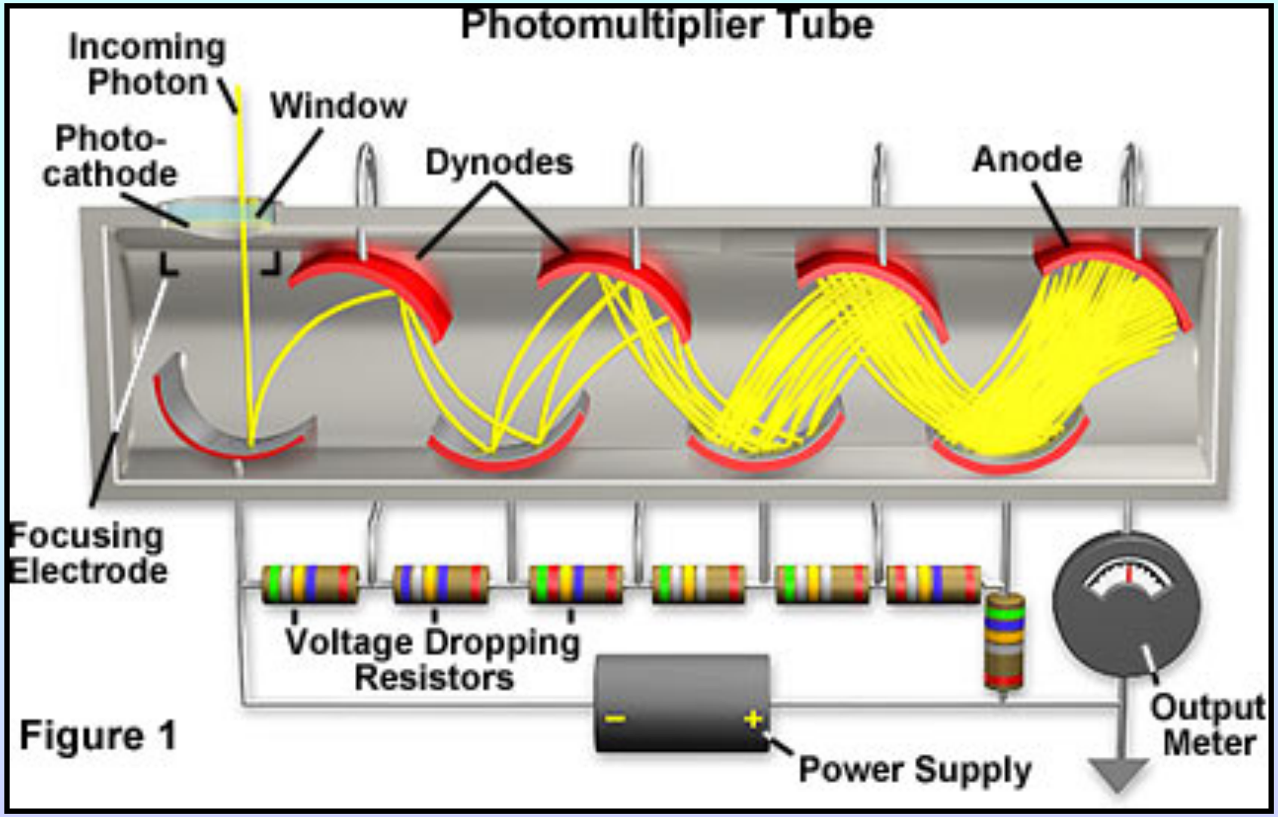


Figure 1

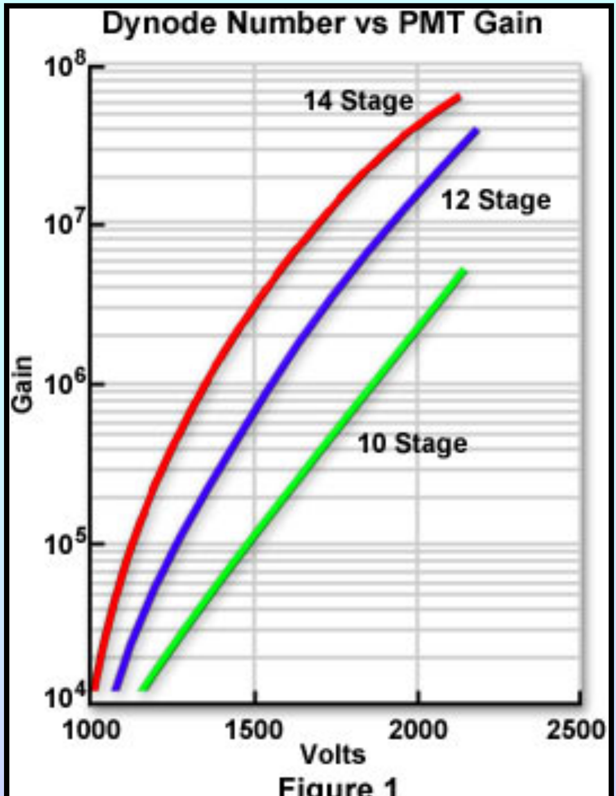
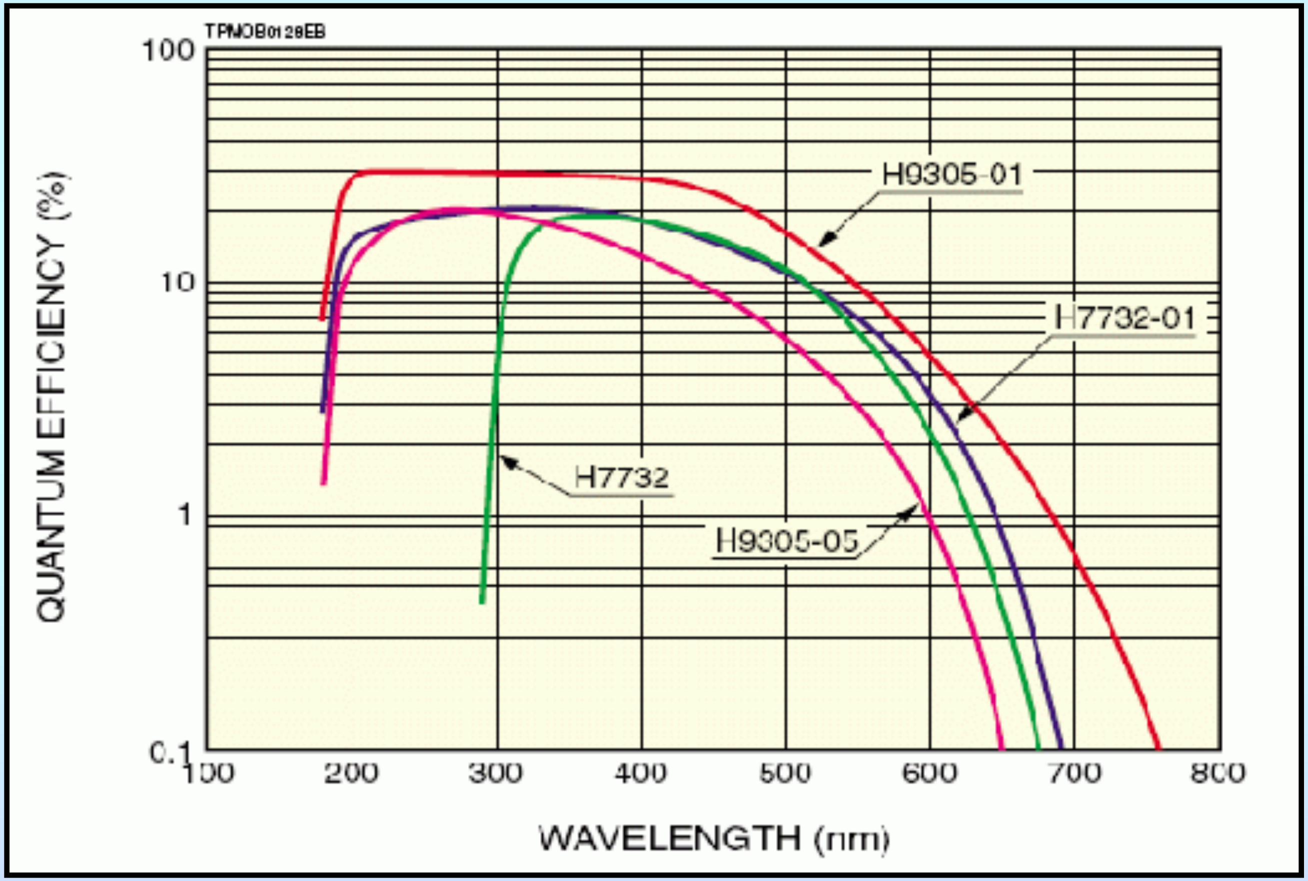


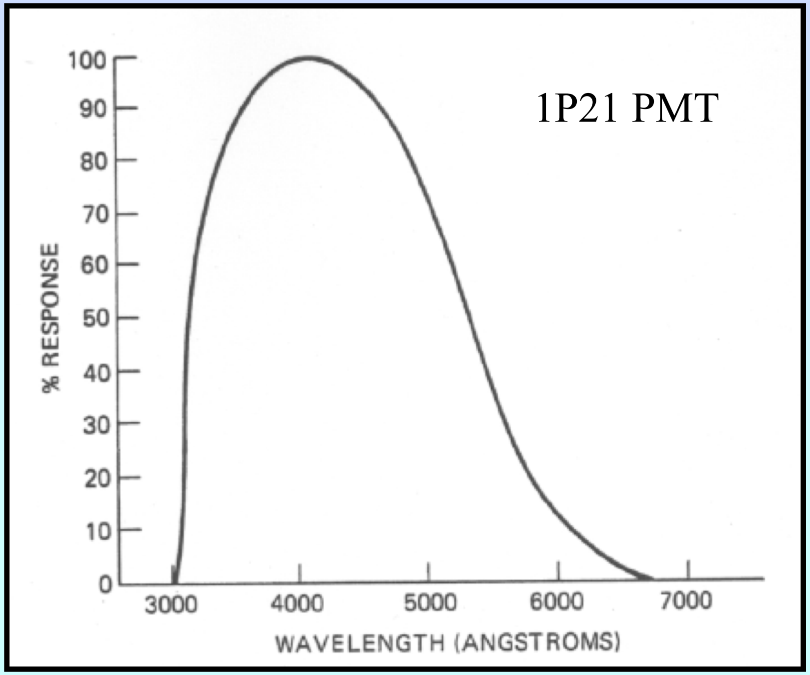
Figure 1

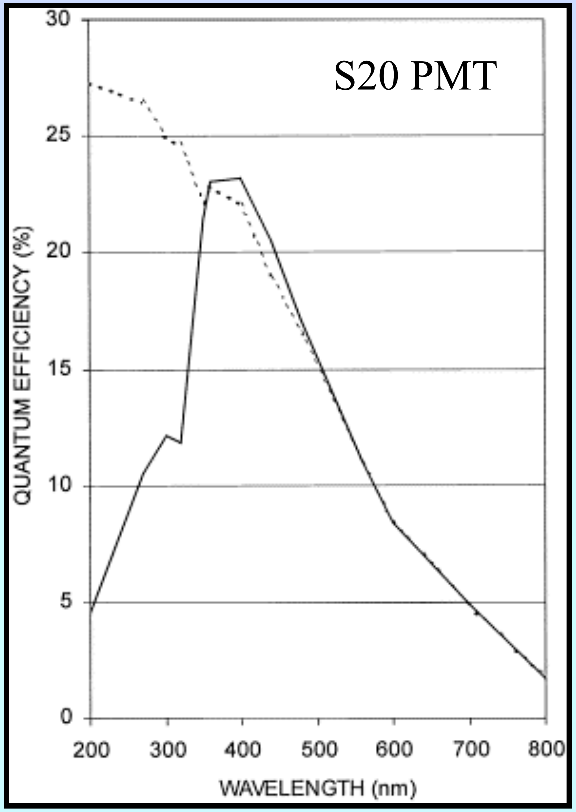
Photomultiplier Tubes

The most used PMTs employed 1P21 or S-20 photo-cathodes. 1P21s are blue sensitive and have response out to ~ 5500 Å; the S-20s have some response out to ~ 7000 Å.

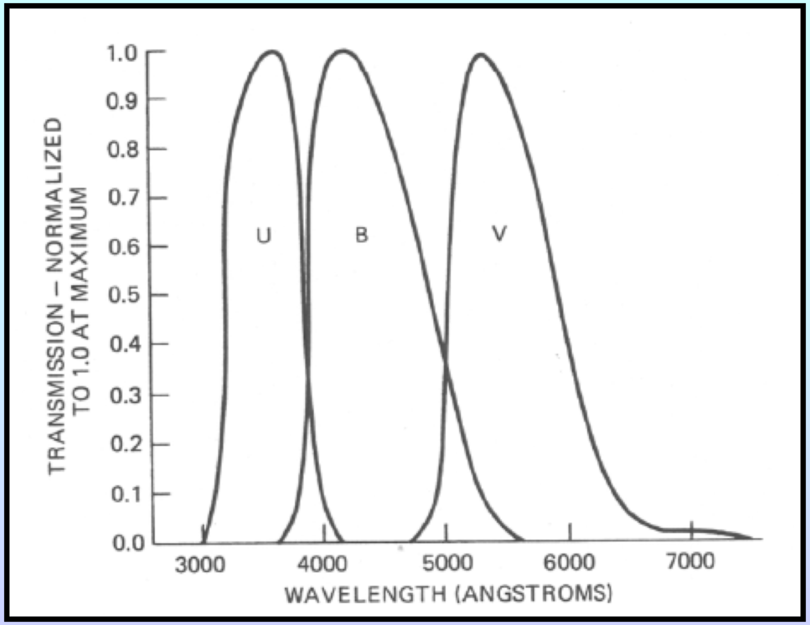


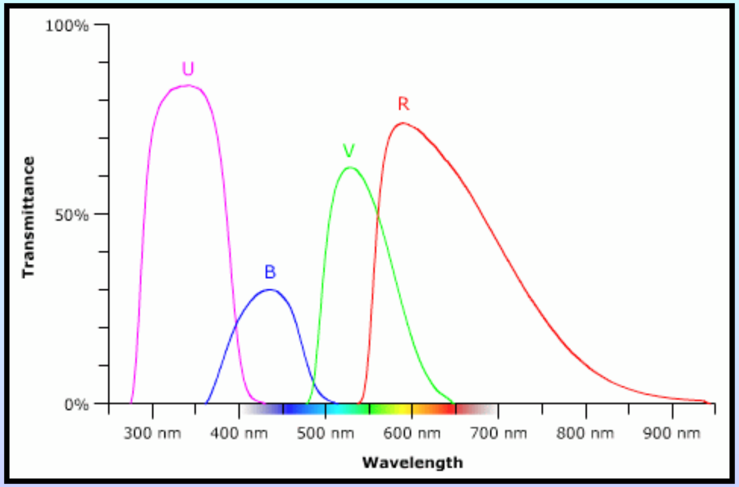
Photomultiplier Tubes and Filter Systems





Traditional photometric systems (such as UBV) are largely defined by (ancient) technology which must be mimicked.





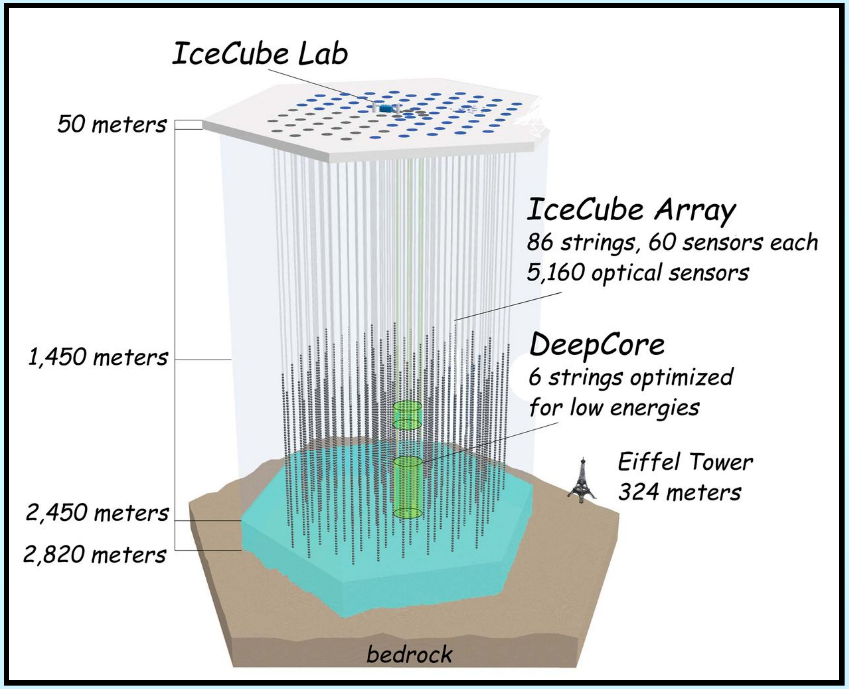
Photomultiplier Tubes

PMT Advantages:

- High gain (so that single photons can be observed easily!)
- Linear detector
- Very little noise
- Photons are time-tagged

PMT Disadvantages:

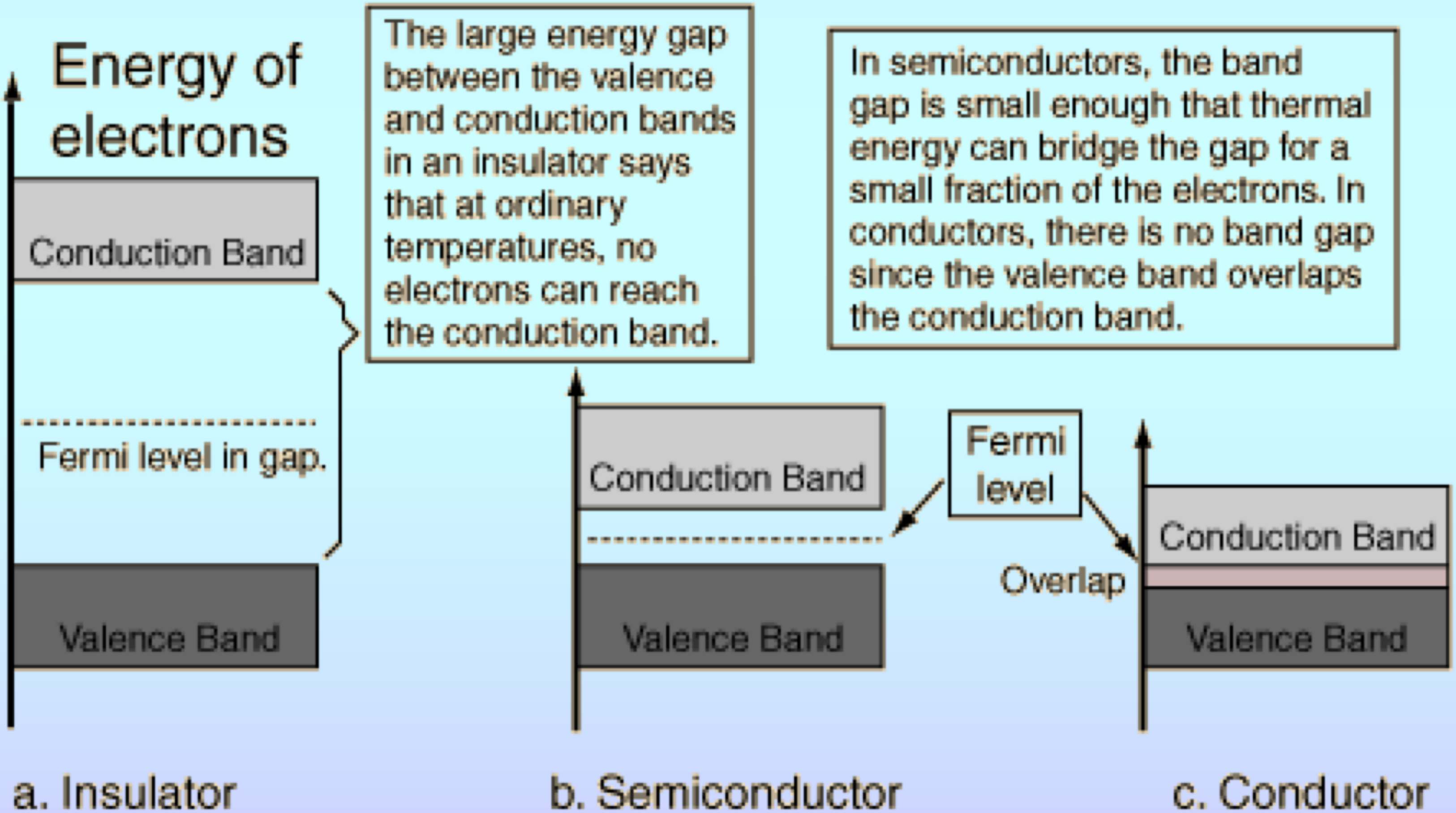
- No spatial resolution (i.e., single pixel detector)
- Response decays in the red (and no IR response)
- Can only detect one photon at a time (requires a correction for coincidence)
- Relatively low quantum efficiency (~13% up to ~25%)



The Ice Cube neutrino experiment at the south pole uses thousands of PMTs buried in the ice.

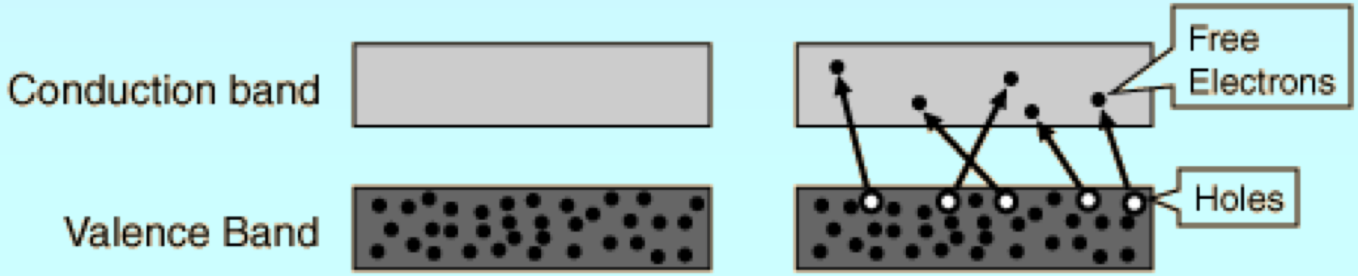
Semiconductors

Since the mid-1980s, semiconductors have been the detector of choice for astronomers. These are solids where the gap between where the electrons normally reside (the "valence" band) is ~ kT below where the electrons can move freely (the "conduction" band).



Semiconductors

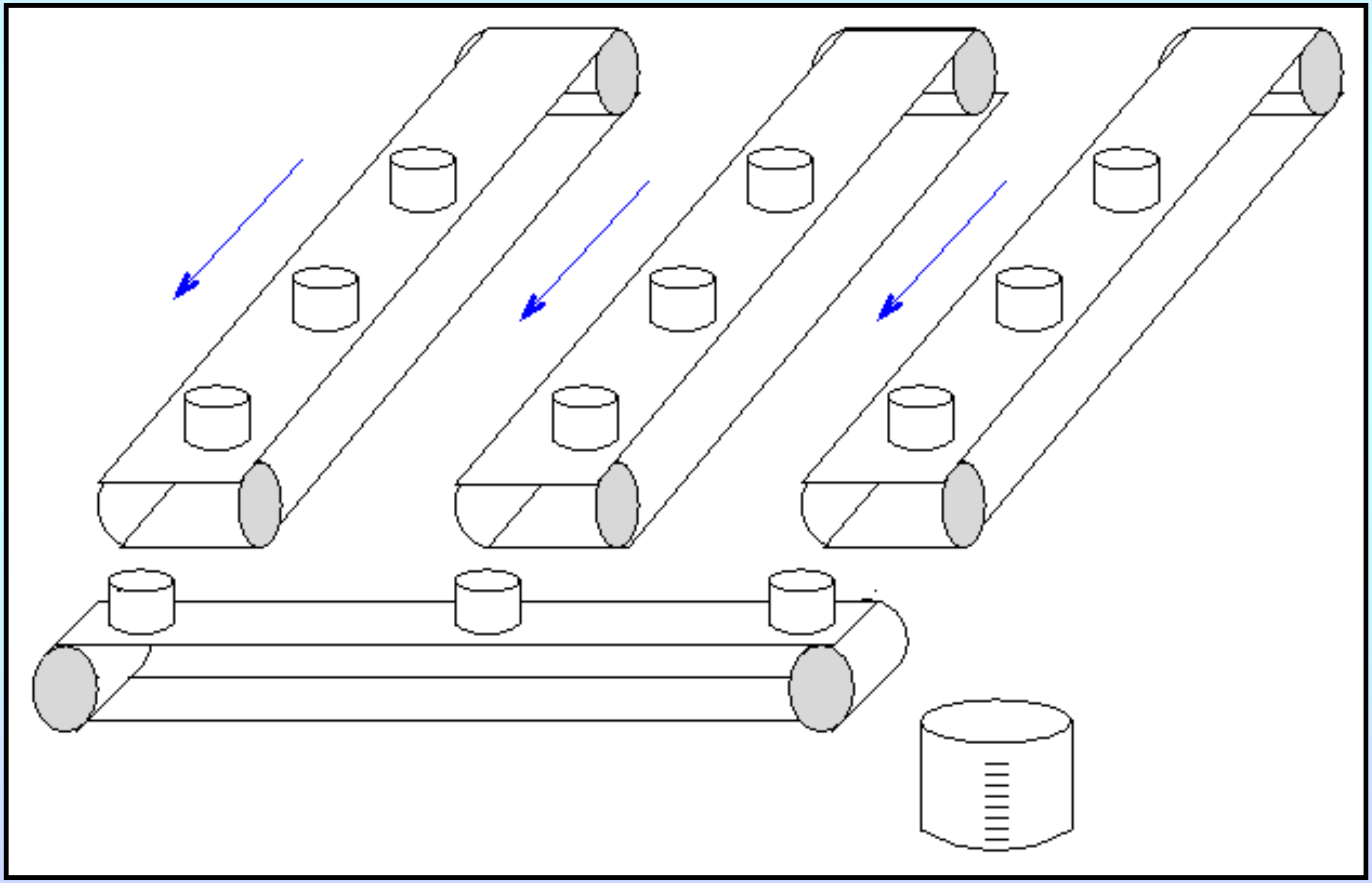
The key to semiconductors is to have a few (~ 4) electrons in the valence band and a small Fermi gap. When a photon with $E > E_g$ hits an electron, it creates a hole in the valence band. A current can then cause electrons to move into the holes.



Type	E_g (eV)	λ_{cutoff}	Detector
GaAs	1.43 eV	0.9 μm	PMT
Si	1.11 eV	1.1 μm	CCD
Ge	0.67 eV	1.8 μm	IR array
InSb	0.23 eV	5.4 μm	IR array

CCDs

CCDs are the detectors of choice for modern optical/near-IR astronomy. Holes accumulate in the valence band and read-out by applying a charge sequentially. It works like a bucket brigade.



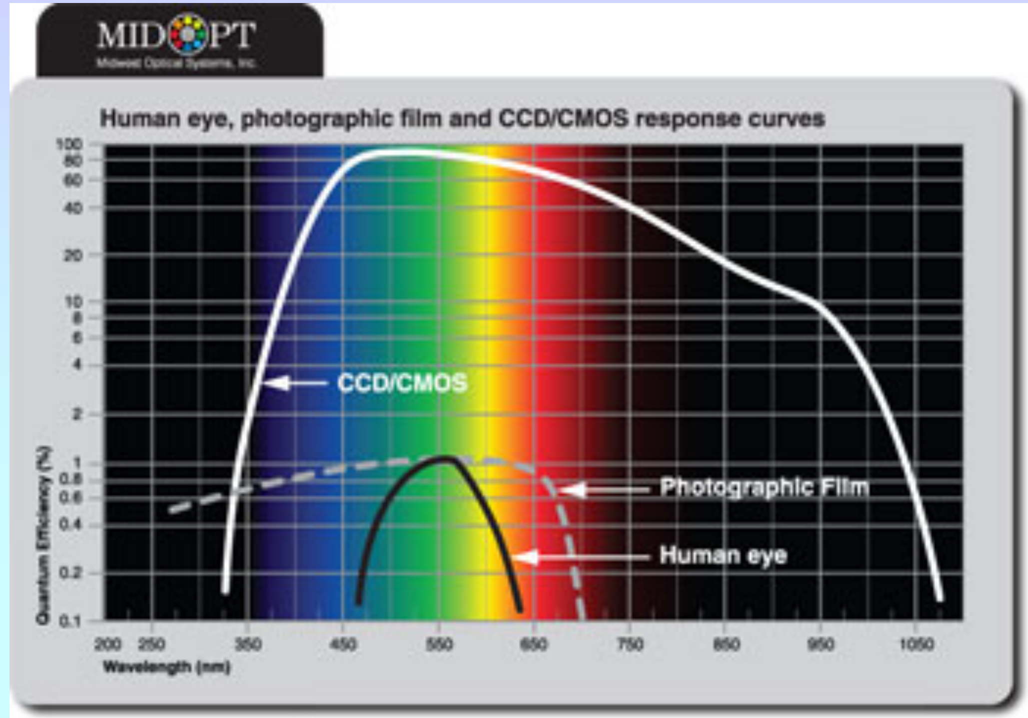
CCDs

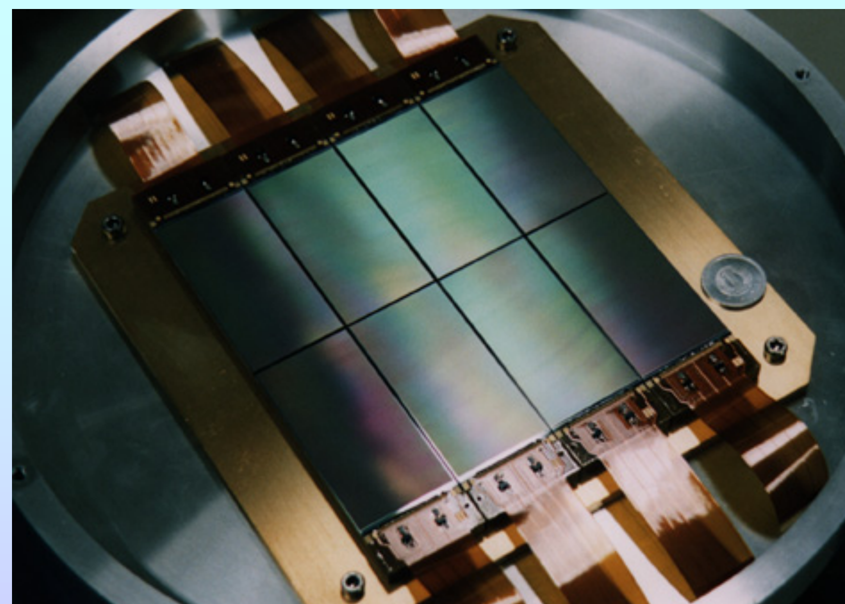
CCD Advantages:

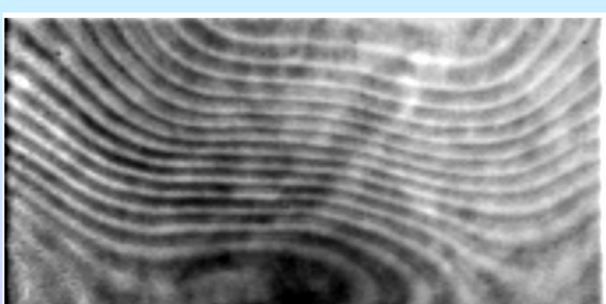
- High(est) throughput
- Large format (up to 4k x 4k)
- Linear detector

CCD Disadvantages:

- Readout noise, dark current
- Not as efficient in the blue
- Limited well depth
- Limited time resolution
- Possible charge transfer issues
- Possible CCD defects, fringing

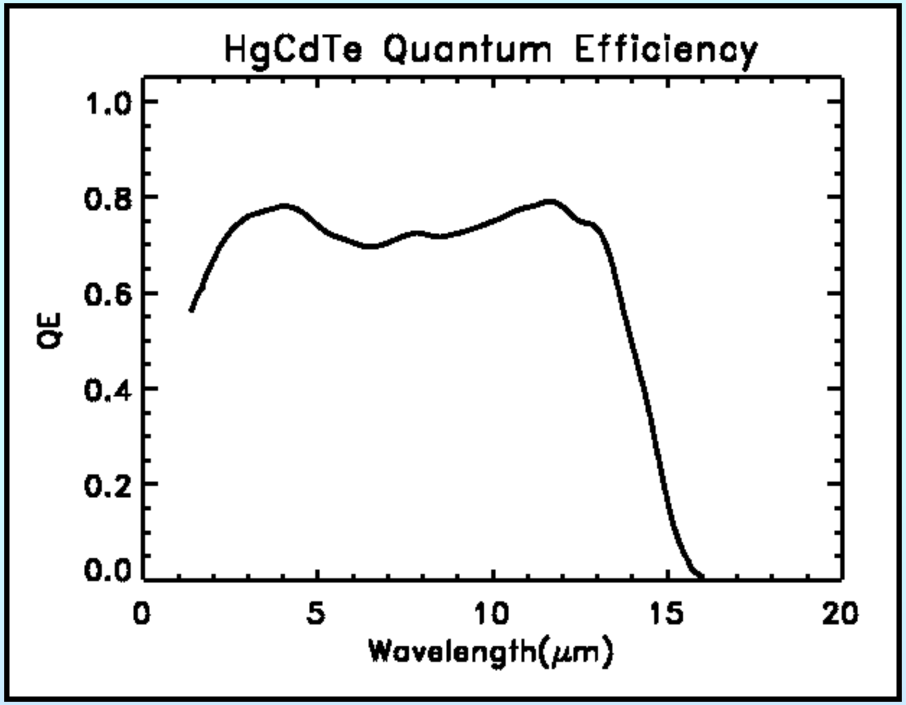


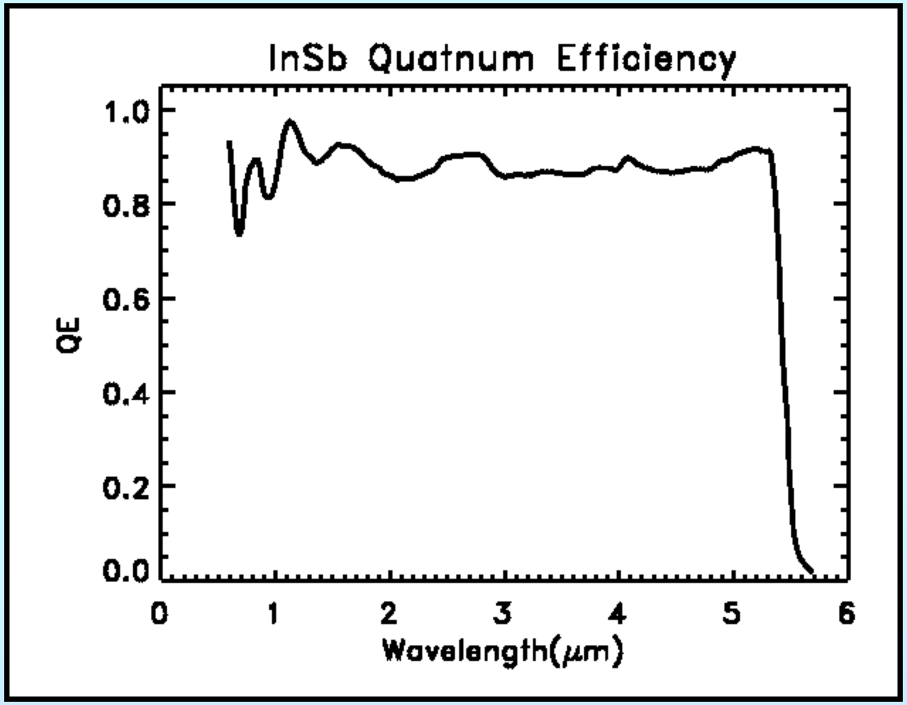




IR Detectors

The Fermi Gap of Si is 1.1 eV, so CCDs cannot detect light longward of 1.1 μm. For near-IR measurements, you need other (much more expensive) semiconductors. The current generation of IR arrays are made of either HgCdTe and InSb and get as large as 2k x 2k.

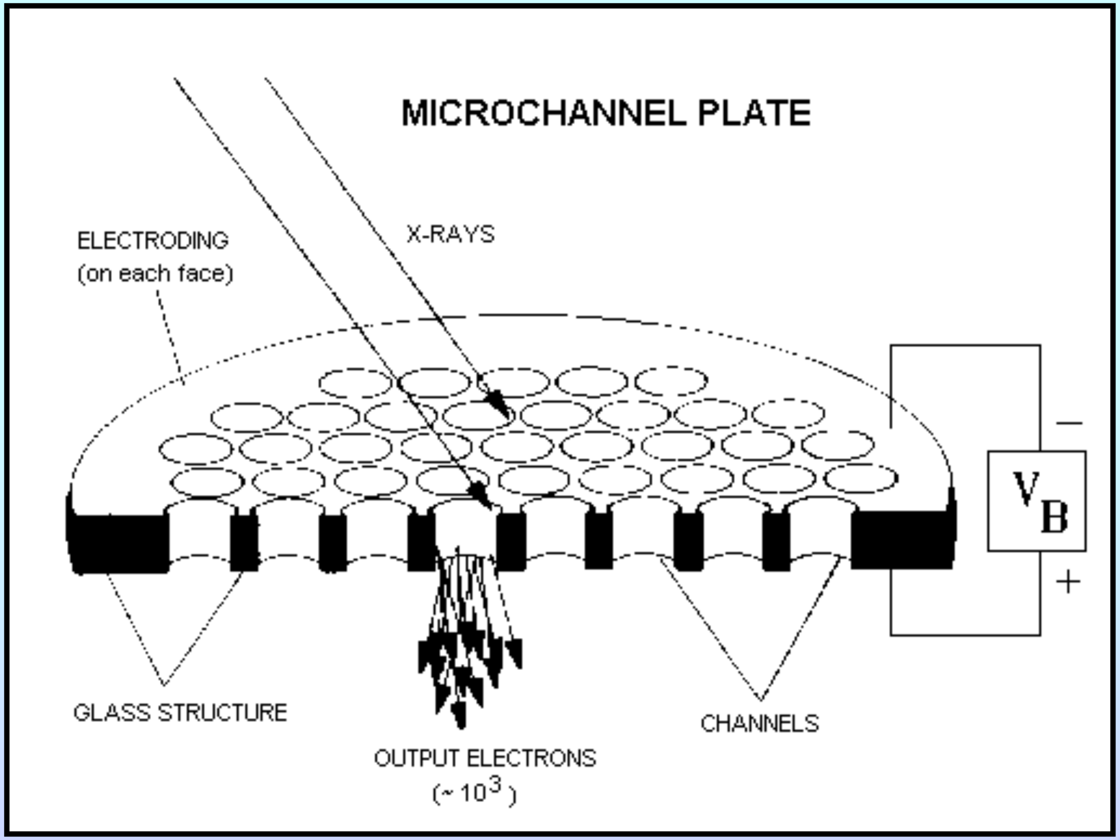


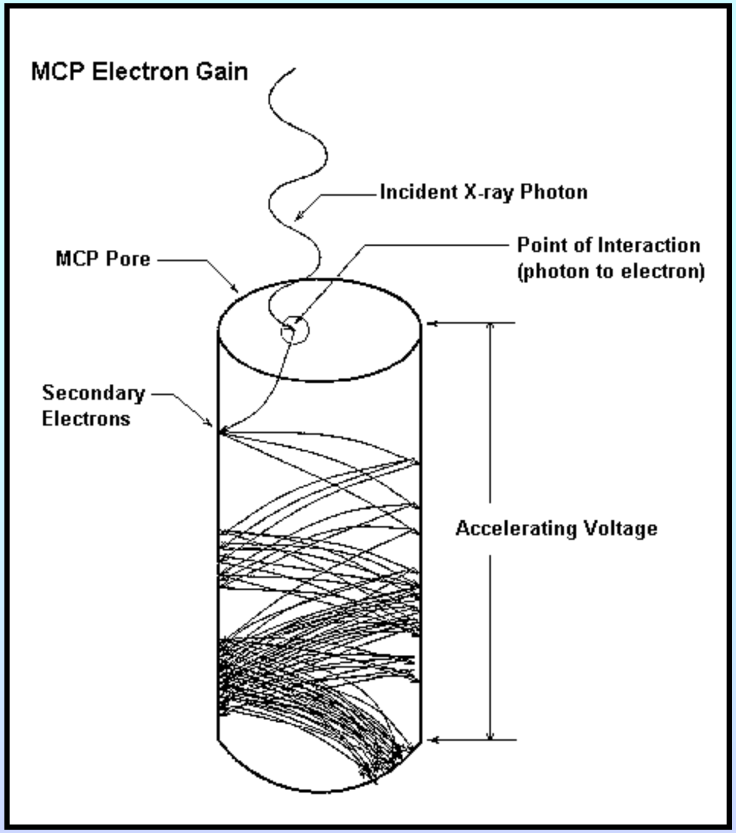


The read noise of IR devices can be large, but this can be improved by moving the charge back and forth to perform multiple reads.

Microchannel Plates

For UV and X-ray detections, Microchannel plates are popular. MCPs consist of a series of small tubes on a slab of photocathode material. The device acts like a collection of PMTs; after the series of dynodes, the electron cloud hits a phosphor, whose light is detected by a CCD. The centroid of the light gives the position of the photon.

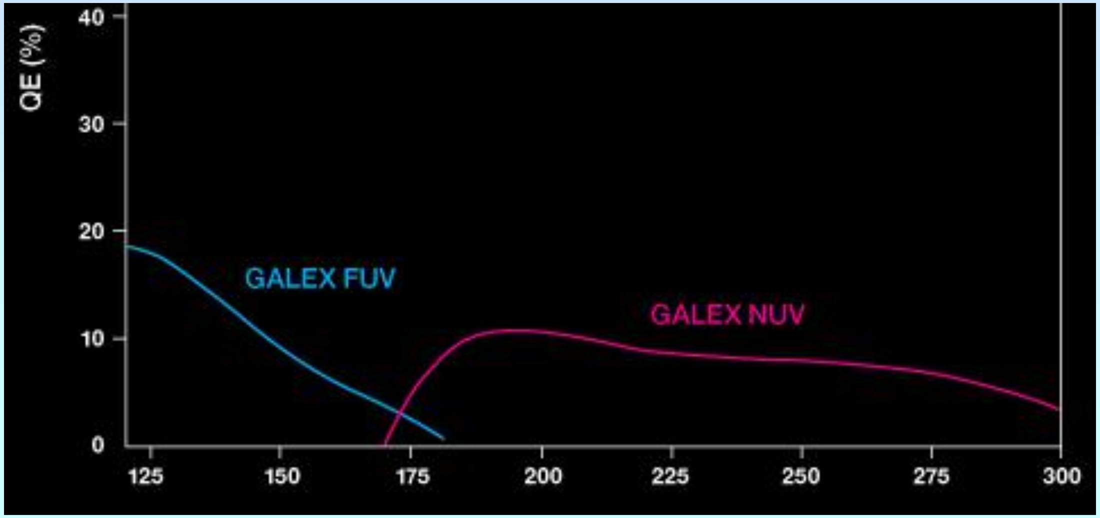




Microchannel Plates

MCP Advantages:

- Single photons detected
- Very little noise
- Linear detector
- Photons are time-tagged

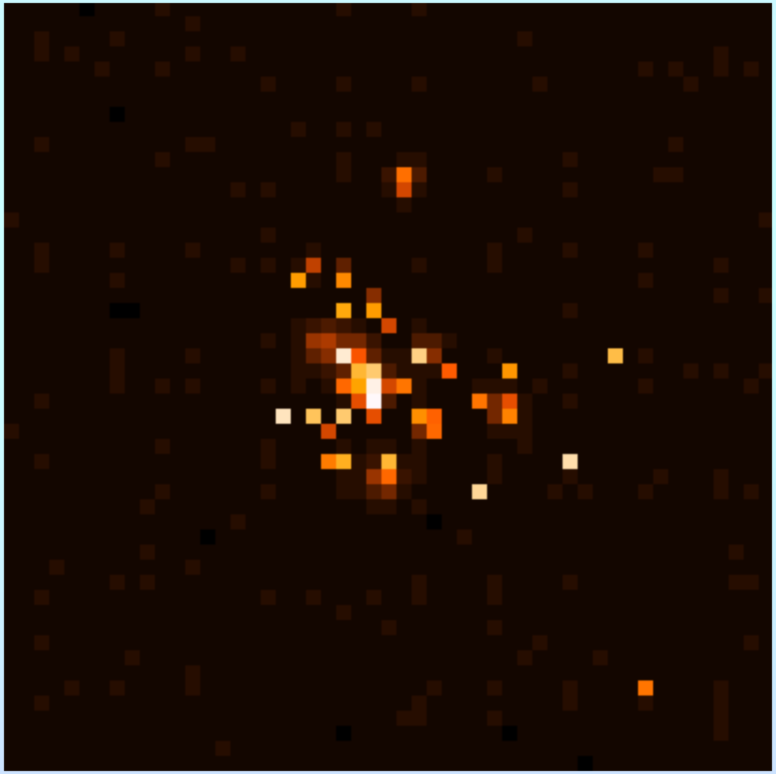


MCP Disadvantages:

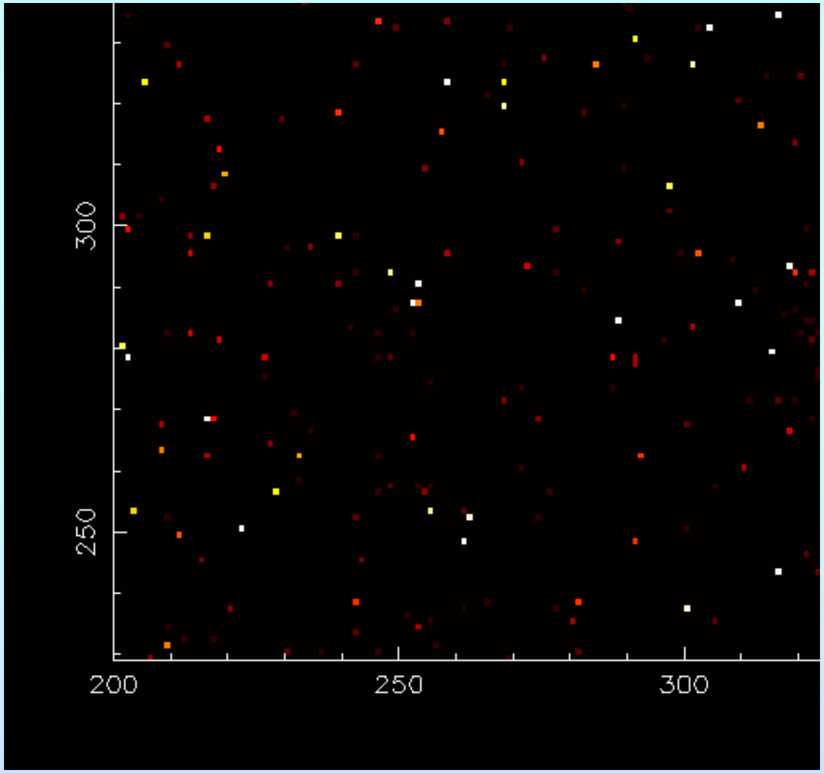
- PMT throughput
- Only one photon can be detected at a time, since the software must find the centroid of the electron cloud. If two photons hit at once, one is lost. Thus one must apply a “co-incidence” or “dead-time” correction.
- Too high a count rate can damage the detector

CCD for X-ray Detection

It takes w = 3.65 eV to ionize an electron from the lattice of a CCD. Thus, when a X-ray hits a CCD pixel, it will liberate $N_e \sim E_x/w$, electrons. The resulting electron cloud will thus contain information on both the X-ray photon's position and energy.



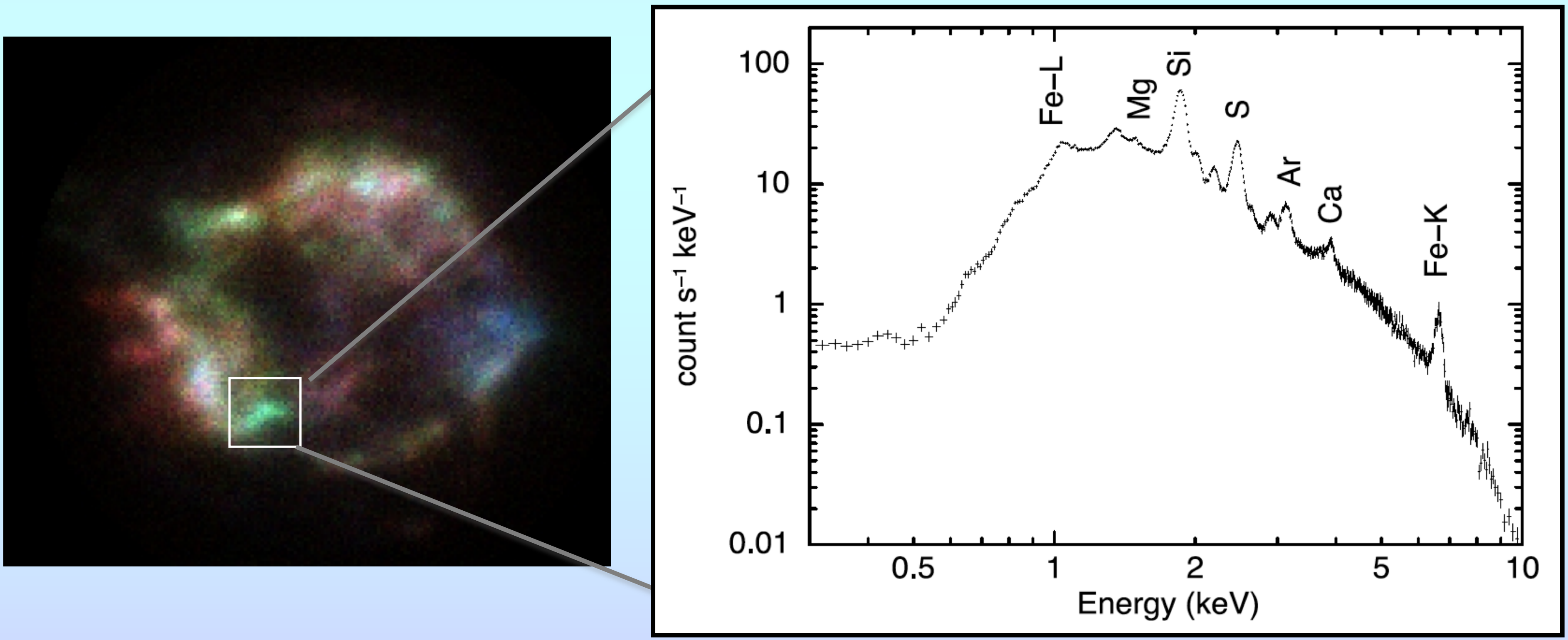
The electron cloud from a single x-ray photon



X-ray detections, with the color denoting the energy

X-ray CCD Spectra

The first-light image of supernova remnant Cas A from the X-ray Telescope (XRT) on Swift. The CCD is read out *very* quickly, to detect individual events. This enables simultaneous imaging, spectroscopy, and “time tagging”.



X-ray CCDs

X-ray CCD Advantages:

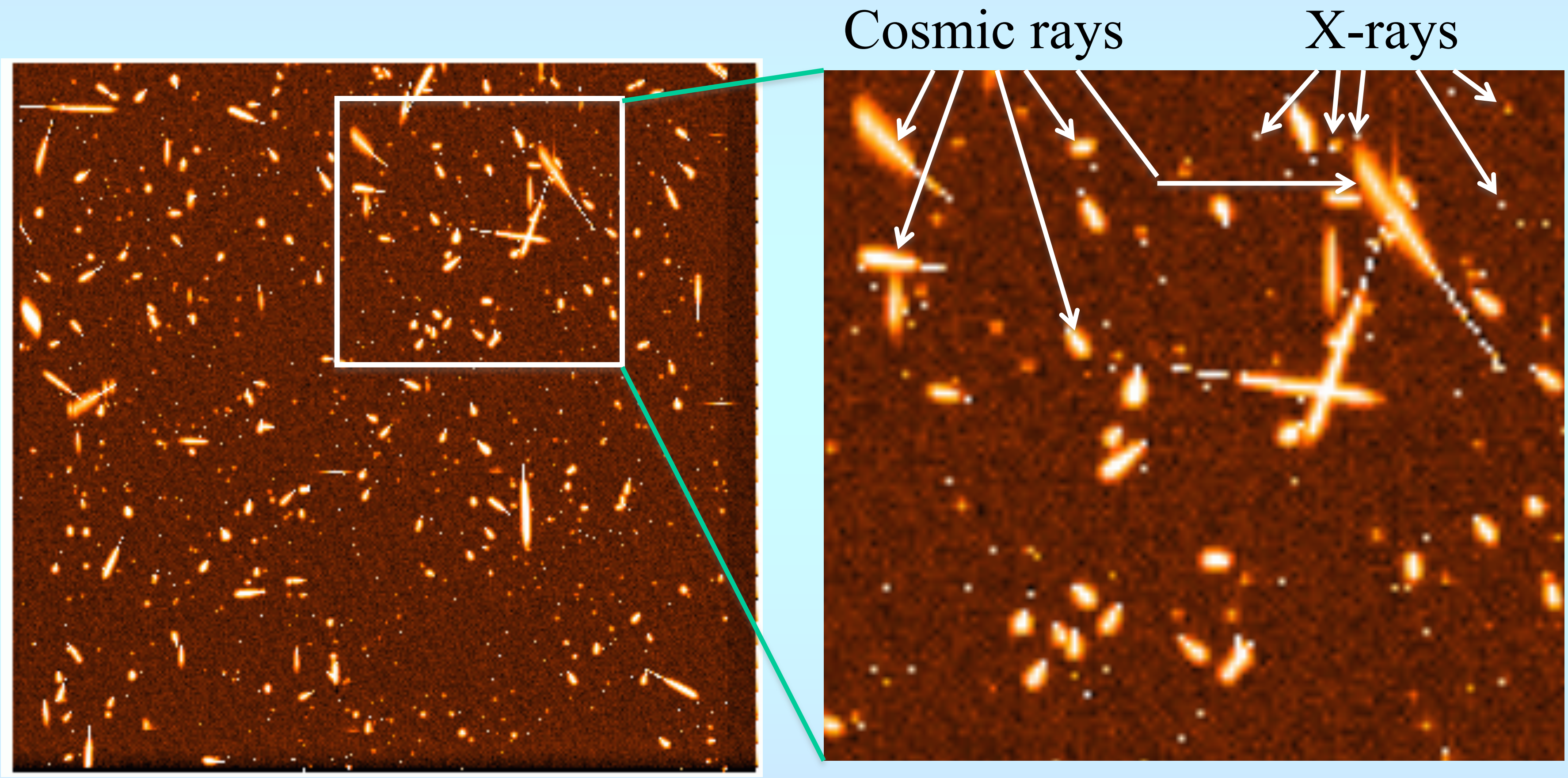
- X-ray photons individually detected
- Simultaneous moderate spectral resolution, with $\sigma_{\mathrm{rms}} \sim (0.1 N_e)^{1/2}$
- Linear detector
- Photons are time-tagged

X-ray CCD Disadvantages:

- Assumes each electron cloud comes from only one photon
- *Lots* of cosmic rays in space
- CCDs susceptible to radiation damage
- “Pile-up” (similar to coincidence loss for MCPs)

X-rays and Cosmic Rays

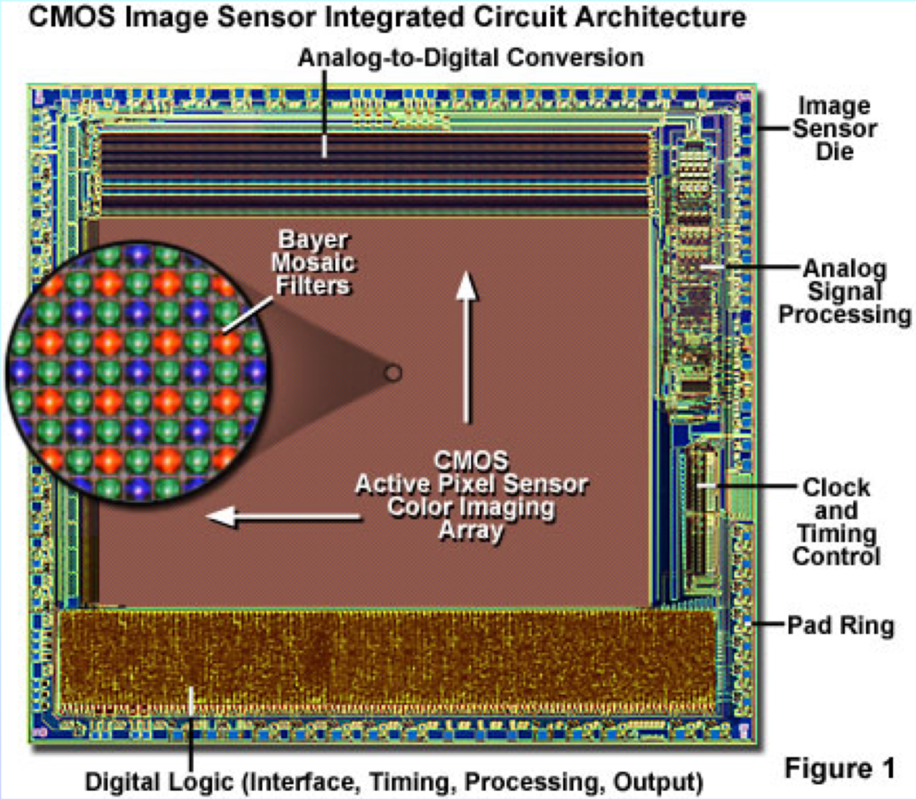
Swift data taken in South Atlantic Anomaly

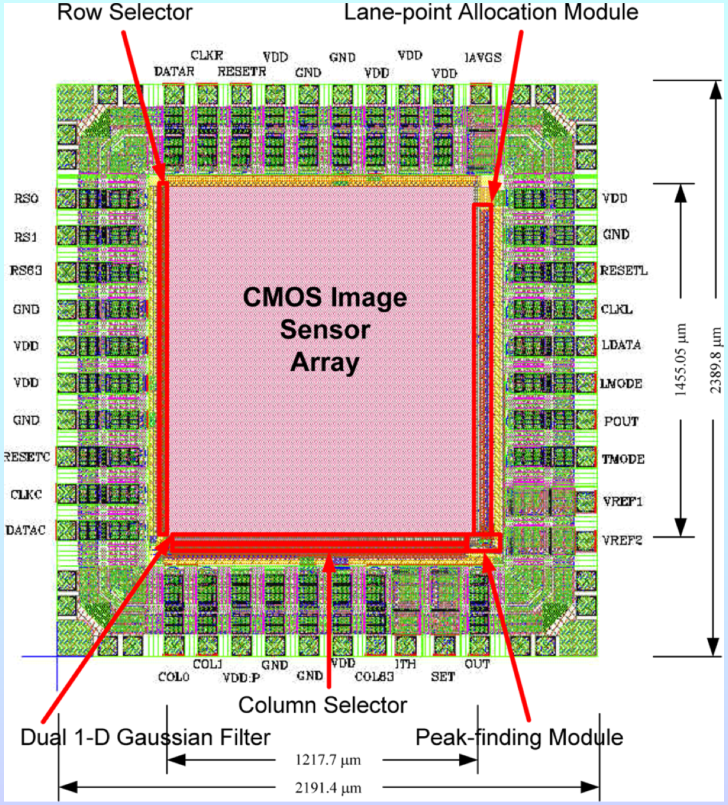


X-rays can be distinguished from cosmic rays using their shape.

CMOS Detectors (the Future?)

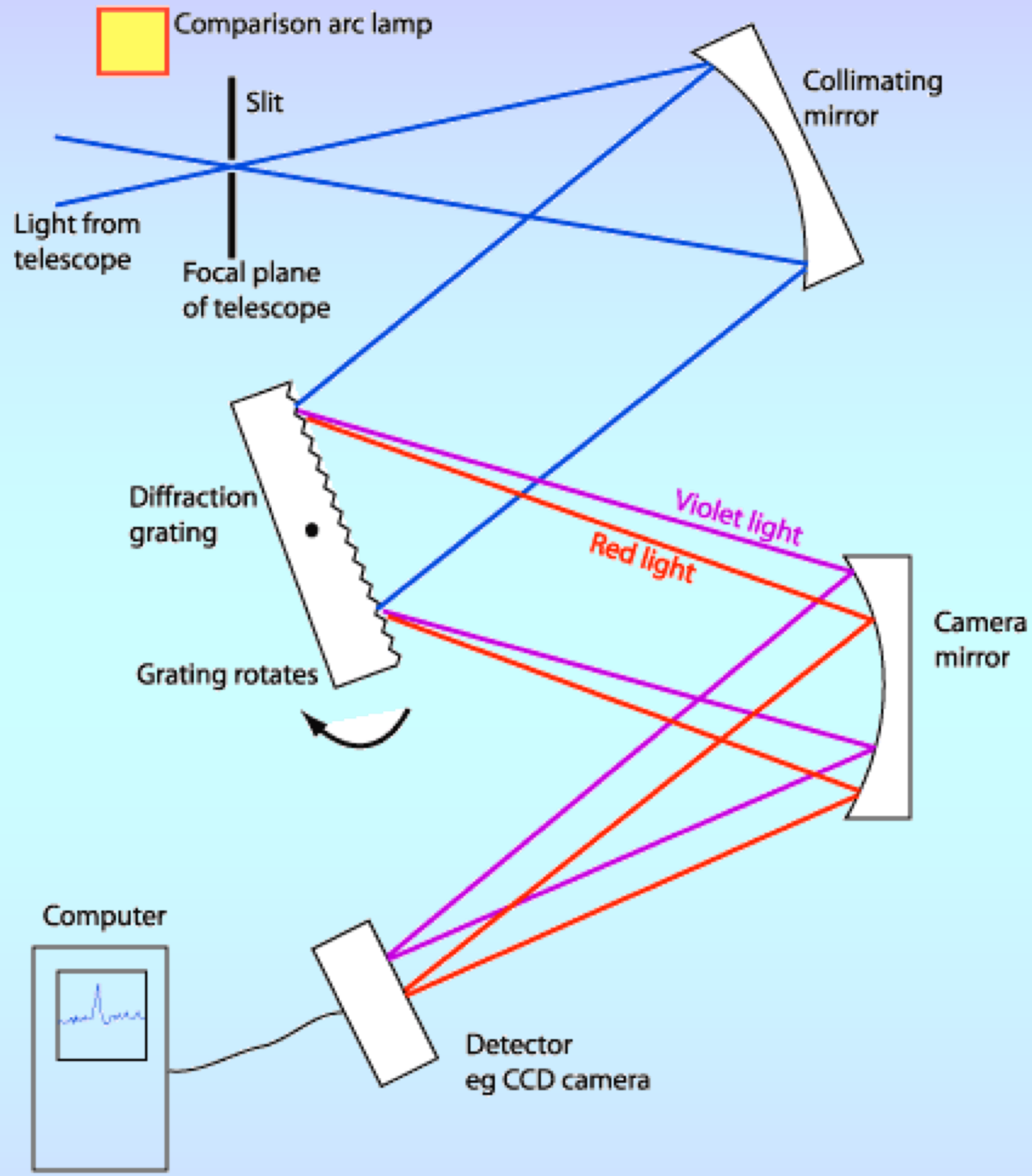
Much work these days (especially for X-ray and IR detectors) involves Complimentary Metal Oxide Semiconductors. In a CMOS detector, each pixel has its own charge-to-voltage converter, amplifier, noise-correction and digitization circuit. The trade-off is uniformity for massively parallel readouts.





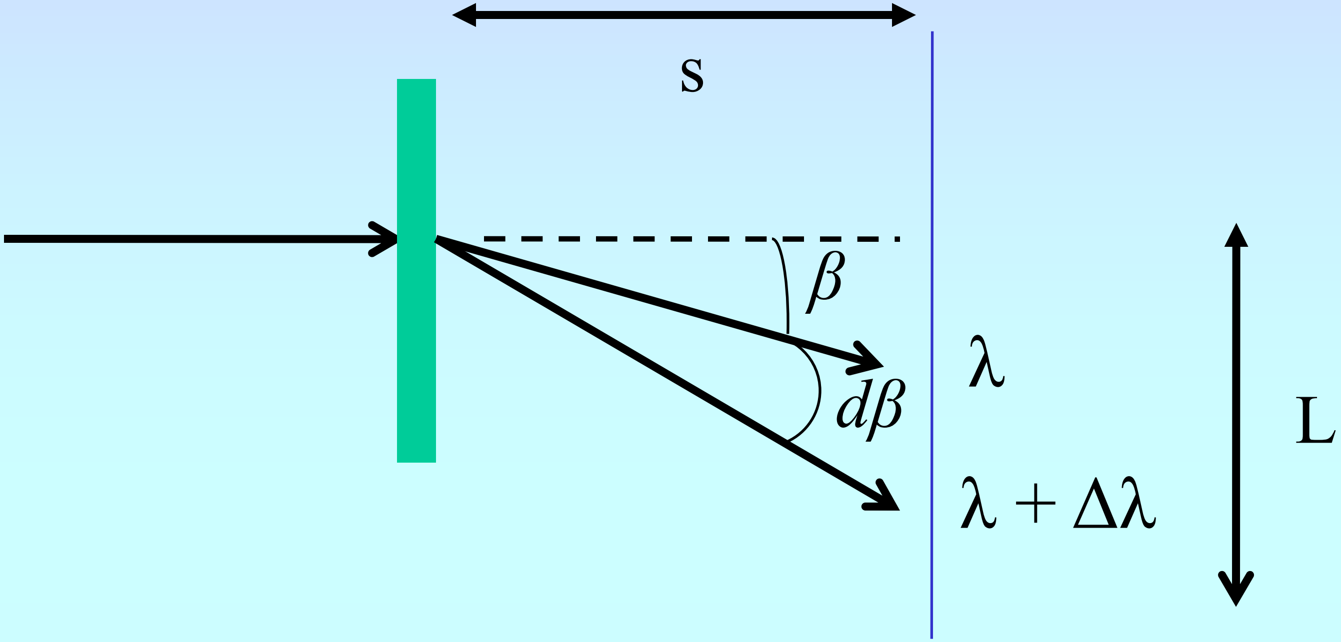
Spectrographs

“ A picture may be worth a thousand words, but a spectrum is worth a thousand pictures. ”



A Schematic Diagram of a Slit Spectrograph

Dispersion



For any spectrograph, the **angular dispersion** is defined as

$\frac{d\beta}{d\lambda}$, while the projected **linear dispersion**, is $\frac{dL}{d\lambda} = s\frac{d\beta}{d\lambda}$

Linear dispersion is usually given in units such as Å/mm or Å/pixel.

Resolving Power

The wavelength "resolution" of an observation is given by

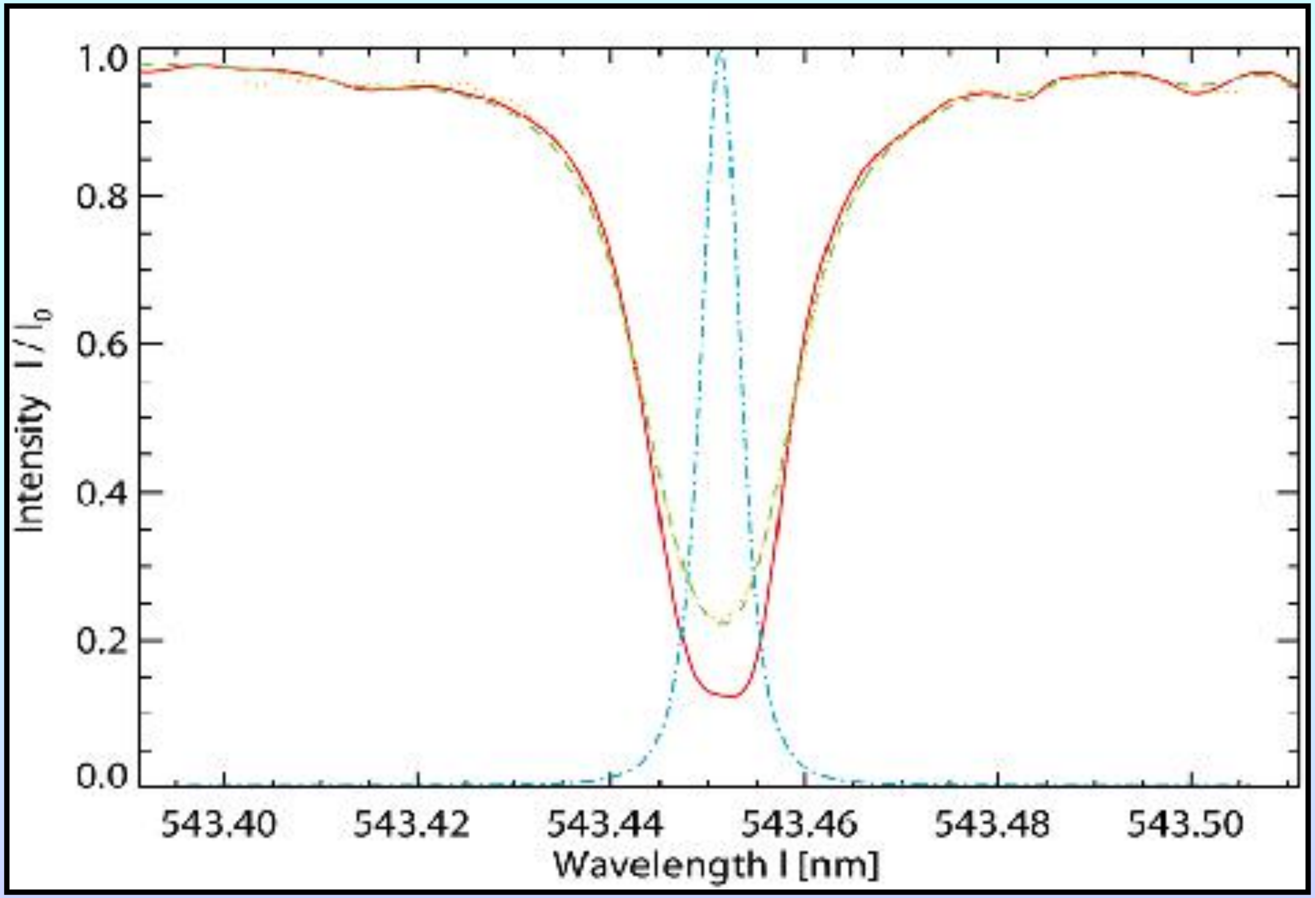
$$R = \frac{\lambda}{\Delta\lambda} = \frac{c}{v}$$

- Broadband Photometry: $R \sim 5\text{-}25$
- Narrowband Photometry: $R \sim 25\text{-}1000$
- Spectroscopy (very general definitions)
  - "Low" resolution: generally $10 \lesssim R \lesssim 2000$
  - "Medium" resolution: $2000 \lesssim R \lesssim 20{,}000$
  - "High" resolution: $20{,}000 \lesssim R \lesssim 1{,}000{,}000$

Note: Resolution is not the same as dispersion, though the two are generally related.

Resolution versus Dispersion

The resolution can be thought of as the width of a delta-function. The dispersion is simply the translation of the linear unit (pixels) to wavelength. For a well-designed in-focus spectrograph, the dispersion is about twice the resolution.



Dispersive Elements: Prisms

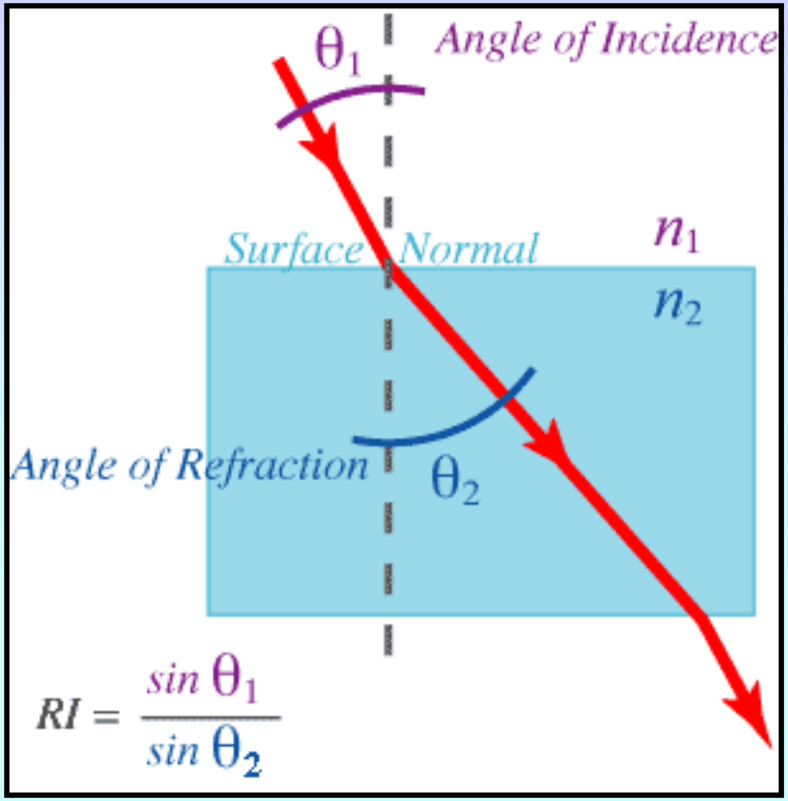
The amount of dispersion produced by a prism is given by Snell's law,

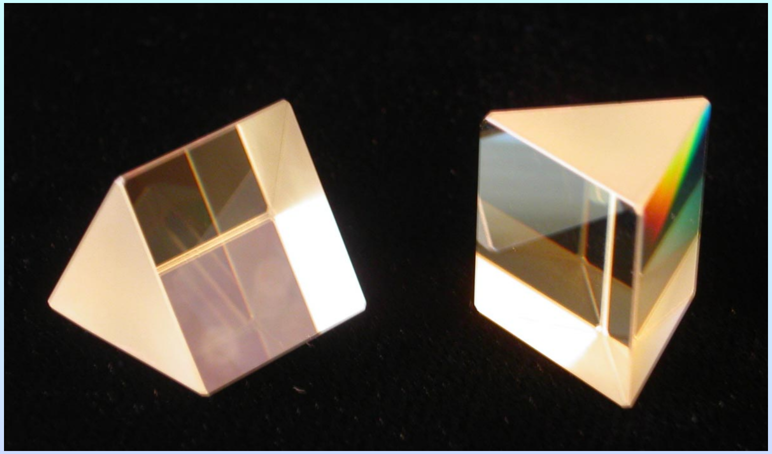
$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Some common indices of refraction are

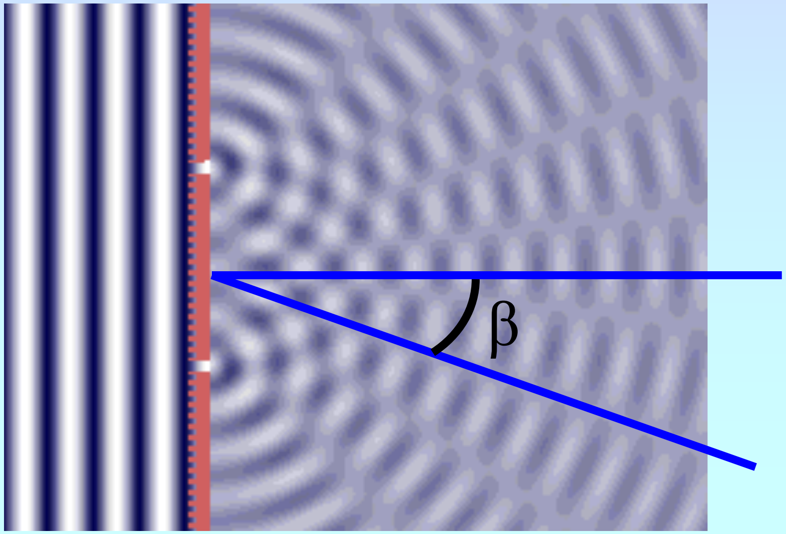
Vacuum	1.00
Air	1.0003
Water	1.33
Ice	1.31
Fused Quartz	1.46
Salt	1.53
Diamond	2.42

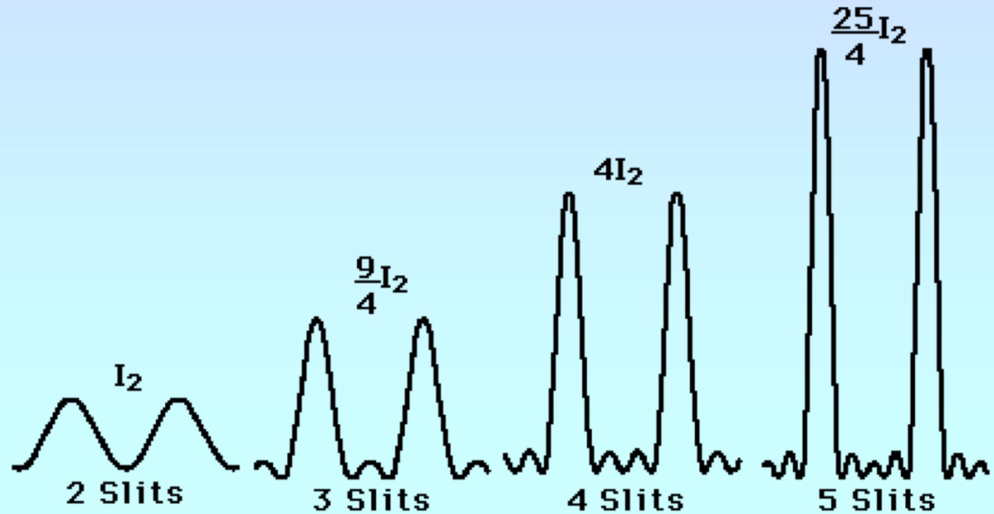
The dispersion of a grism is wavelength-dependent and highly non-linear, making calibrations difficult.





Dispersive Elements: Gratings





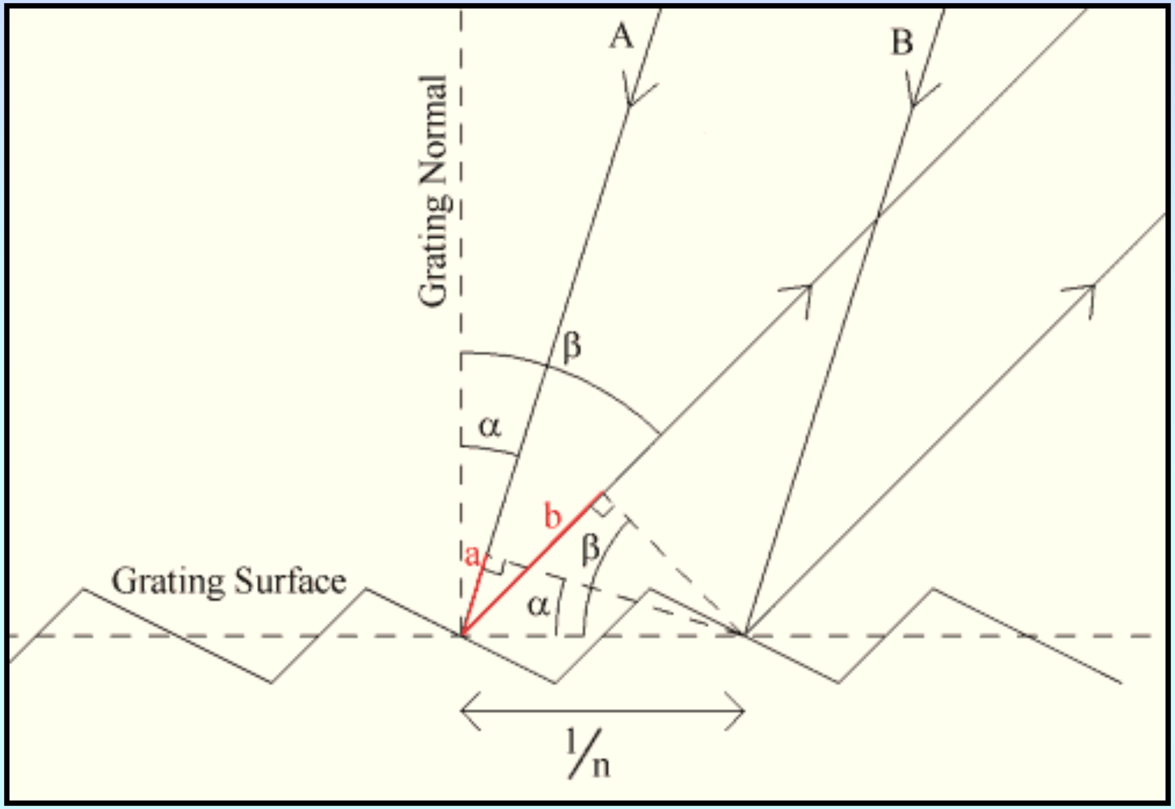
Light entering a set of slits will constructively/destructively interfere. As the number of slits increase, the maxima get brighter, and narrower. This allows the light of different wavelengths to be more dispersed. The relation between the slit separation, *d*, the angle, β, and the wavelength for the **principal maxima** is

$$\boldsymbol{d \sin \beta = m \lambda}$$

where **m** is the spectral order.

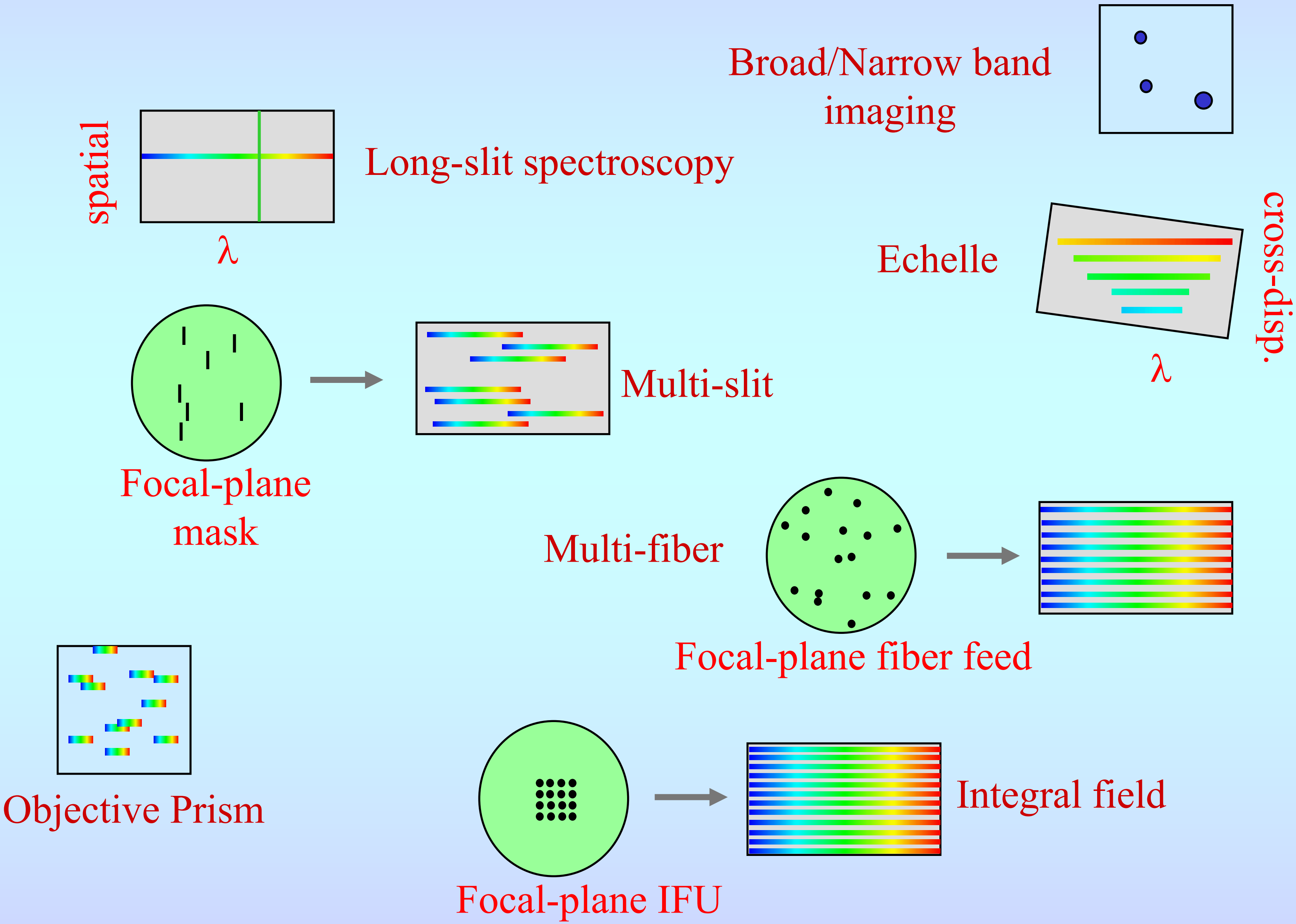
Blazed Gratings

Astronomical gratings are designed to put most of the incoming light into a single order. This is done through the use of a "blaze angle". These ridge angles modify the grating equation to



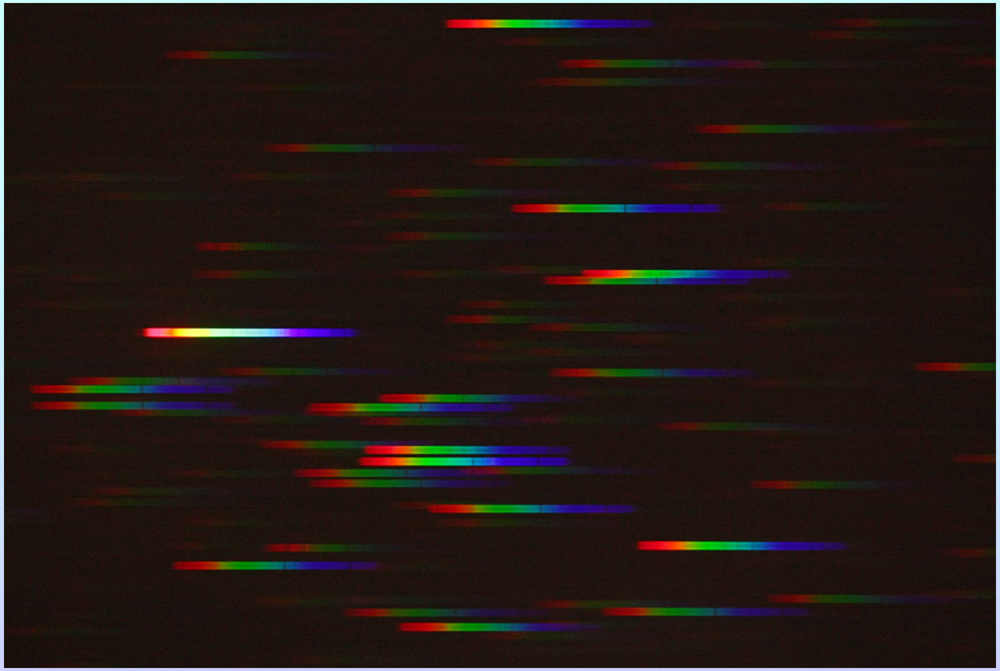
$$d\,(\sin\alpha + \sin\beta) = m\,\lambda$$

Types of Spectroscopy



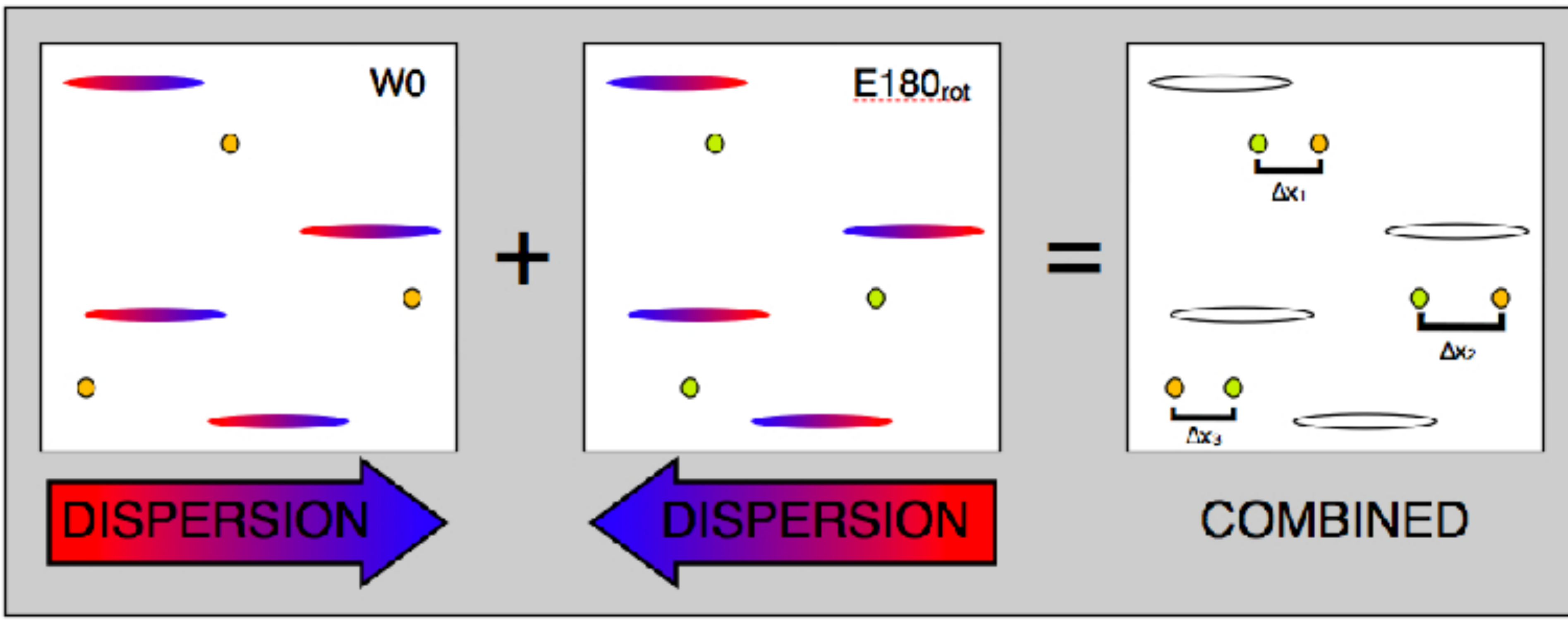
Object Prism or Grism Spectroscopy

If a prism (or grating-prism = grism) is placed in front of the telescope, the result will be a spectrum of everything in the field. Because there are no slits, the background at any position will be relatively high (as the spectrum of every point overlaps those of adjacent points). Thus, "slitless spectroscopy" is usually used for bright objects, or strong emission line sources.



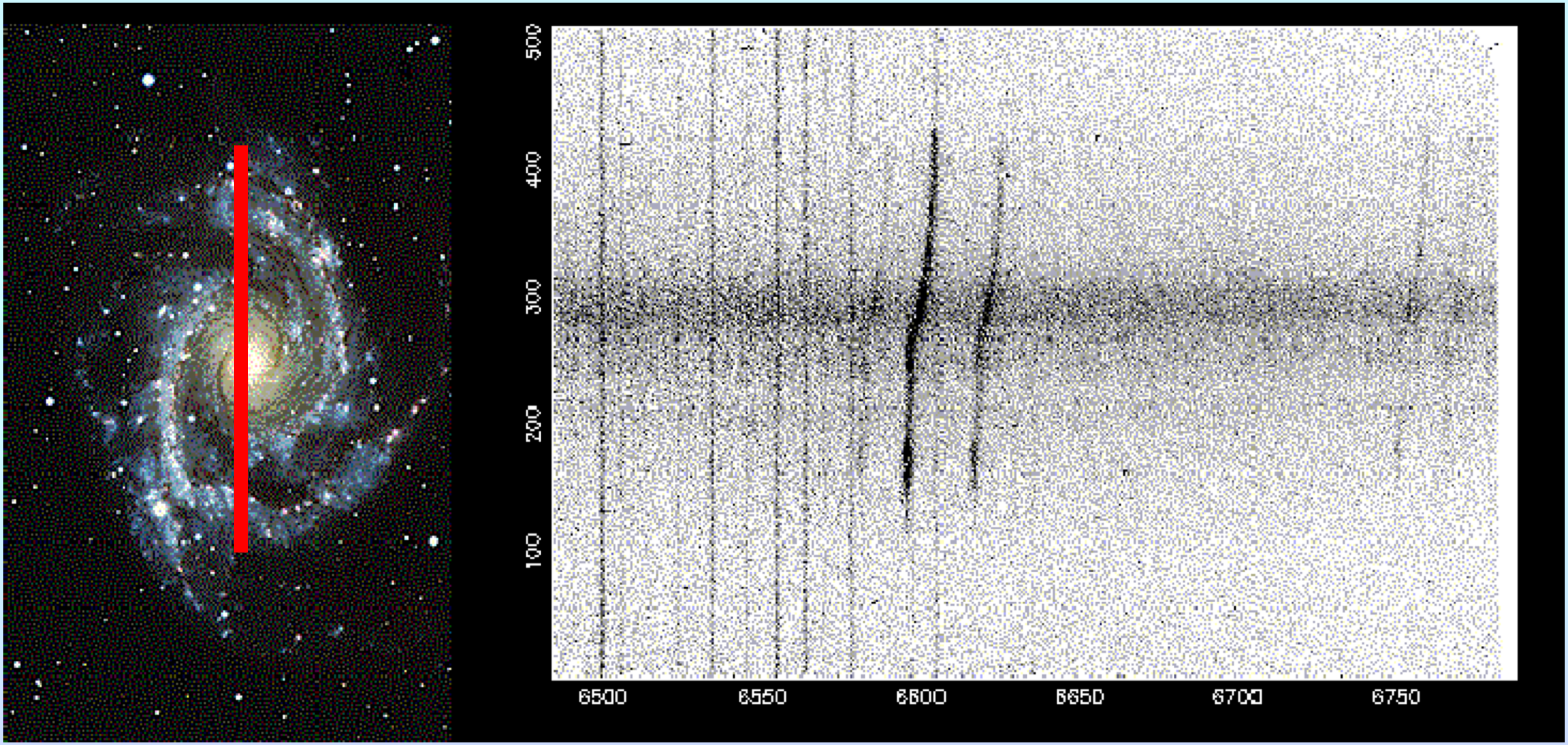
Counter-Dispersed Slitless Spectroscopy

Counter-dispersed slitless spectroscopy is sometimes used to obtain velocities of strong emission-line sources. A filter plus a prism (or grism) produce data over a limited spectral range. Stars will show up as full spectra, while strong emission line sources will appear as points. When the prism is rotated 180˚ , the distance between the "points" depends on wavelength, and is therefore a function of velocity.

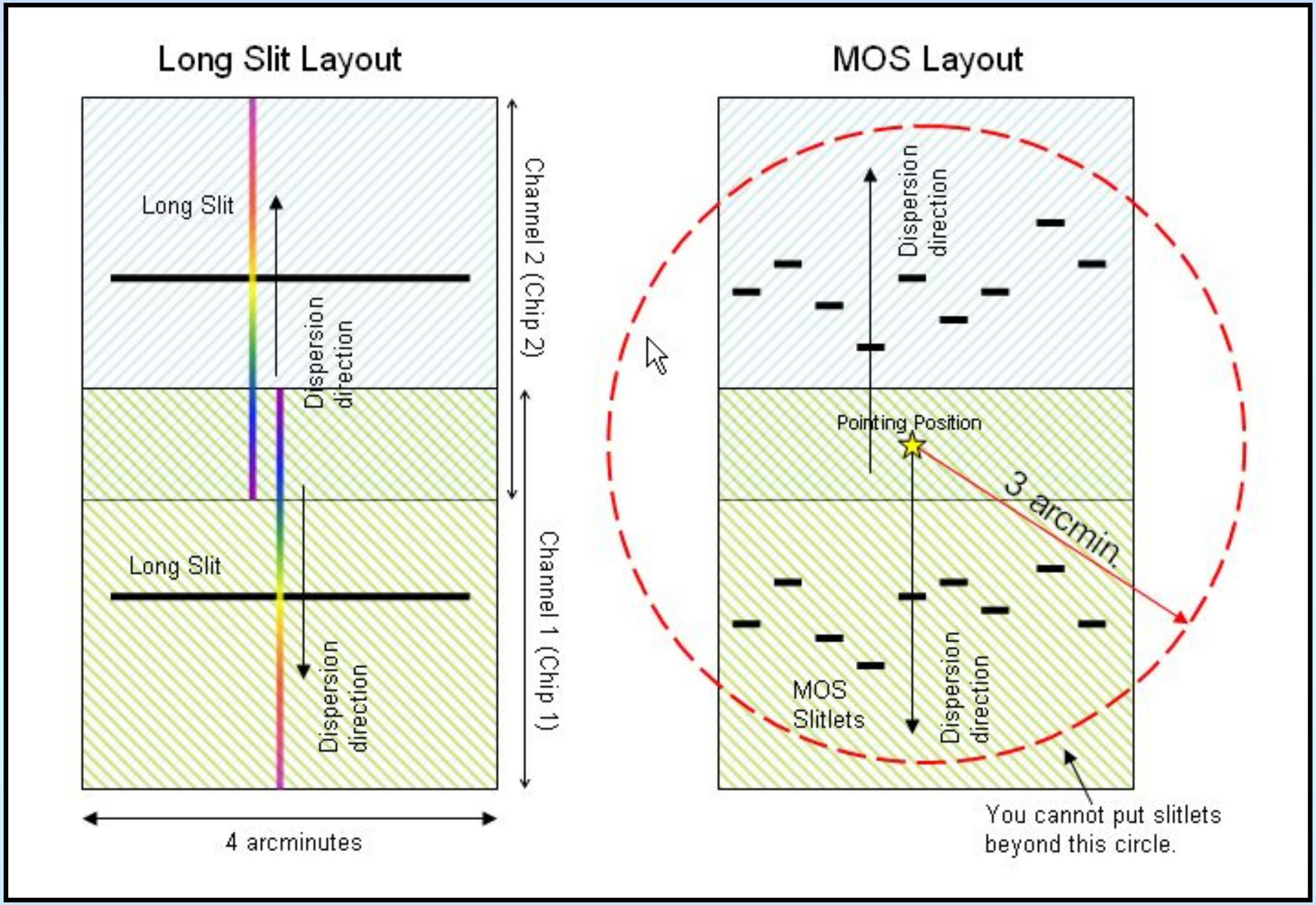


Traditional Long-Slit Spectroscopy

Most astronomical spectroscopy is done through a long slit. This allows for good sky subtraction, but depending on the science, the slit may not be at the parallactic angle.



Long Slit and Multi-Object Slit (MOS) Spectroscopy

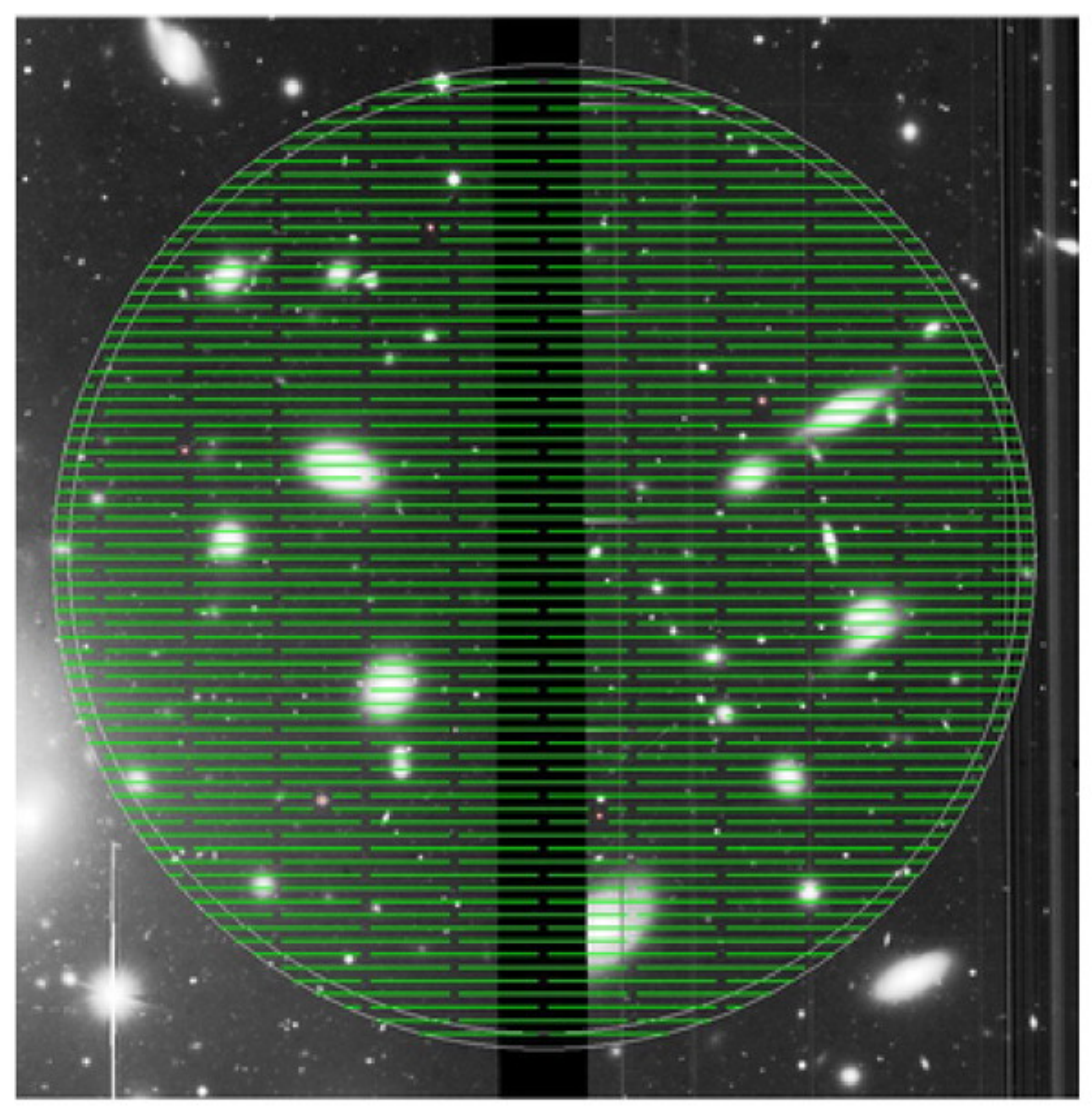


Rather than observe one object with a single slit, you can create a mask to perform multi-object spectroscopy. But the mask must be designed so that the spectra do not overlap.

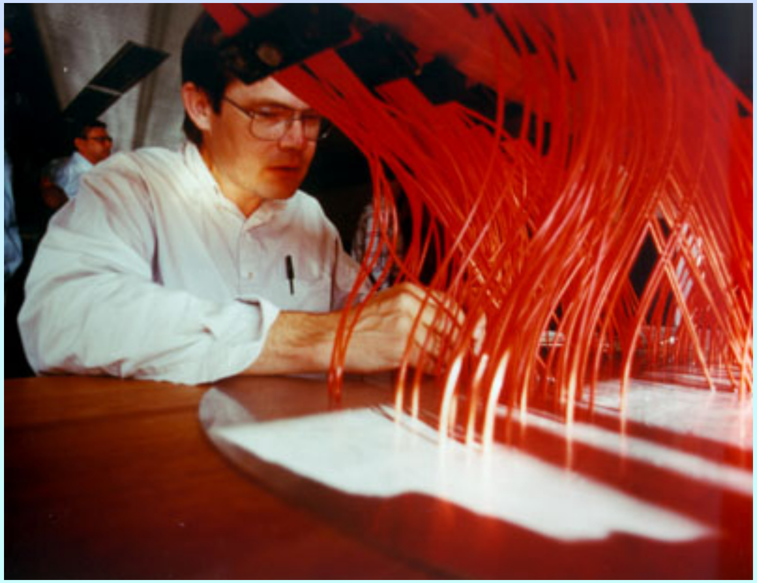
Multislit Imaging Spectroscopy

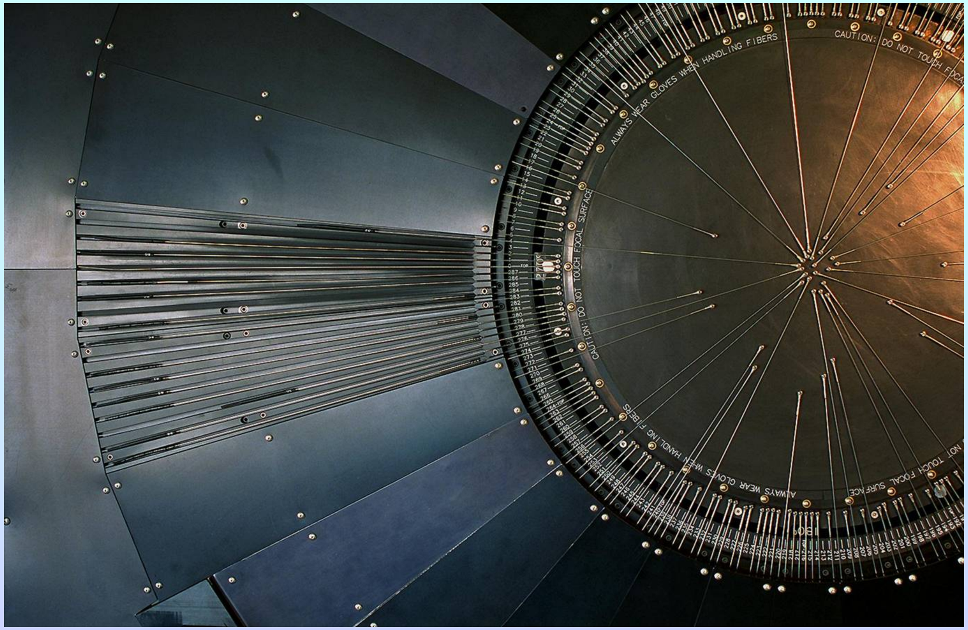
Multislit imaging spectroscopy attempts to combine the depth of the slit spectroscopy with the area of grism spectroscopy. One first restricts the wavelength range with a filter, and then fills the field-of-view with slits (while making sure that the spectra do not overlap).

One can obtain a spectrum of ~ 1/10th the entire area.



Multi-Fiber Spectroscopy

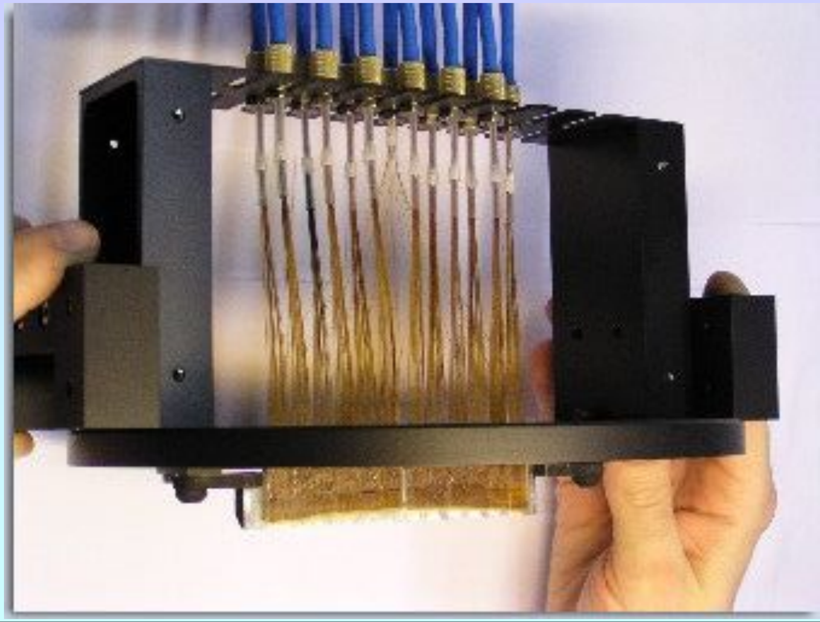


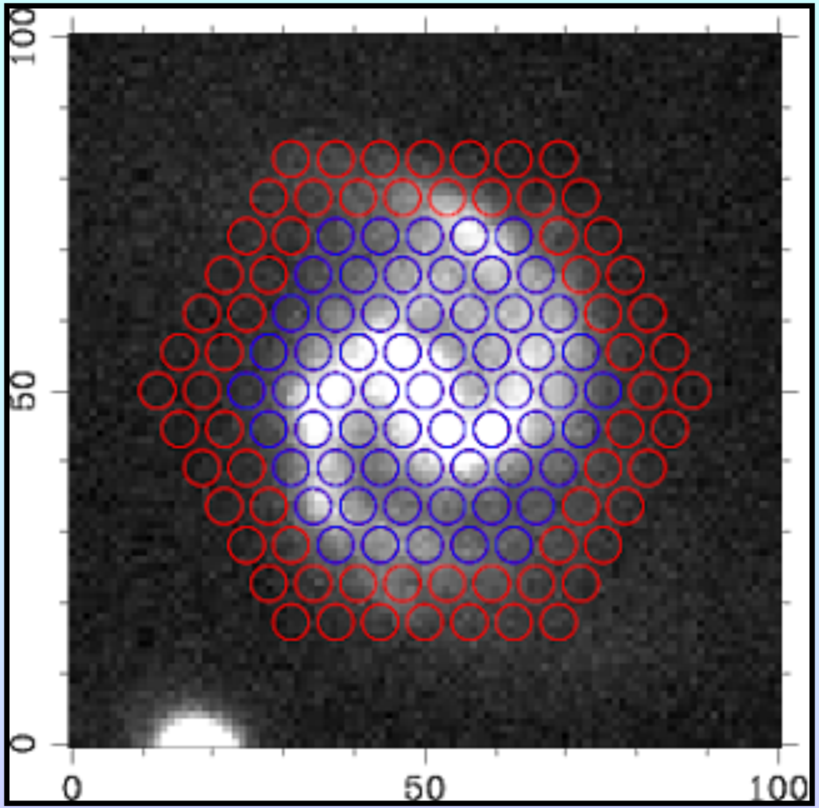


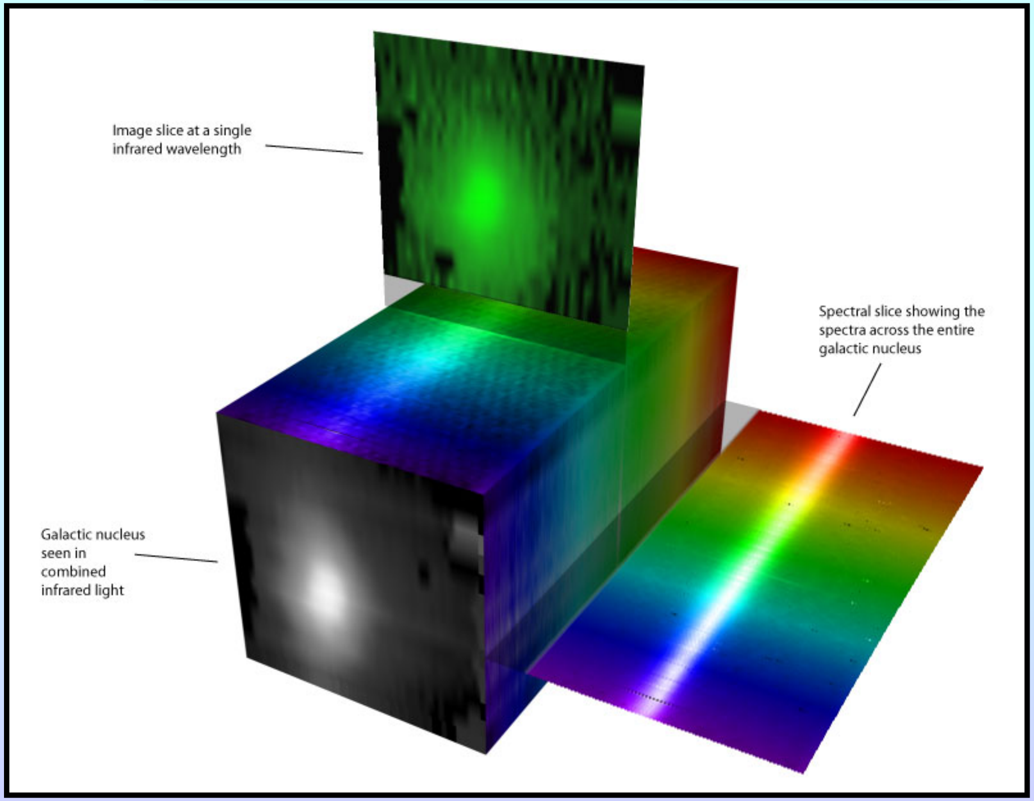
One can use optical fibers to feed a spectrograph and therefore do multi-object spectroscopy. These fibers can be positioned via plugboards or a robotic system.

Integral-Field Unit (IFU) Spectroscopy

By building an array of fibers, one can obtain a spectrum at every location with an object. The result is a data cube.



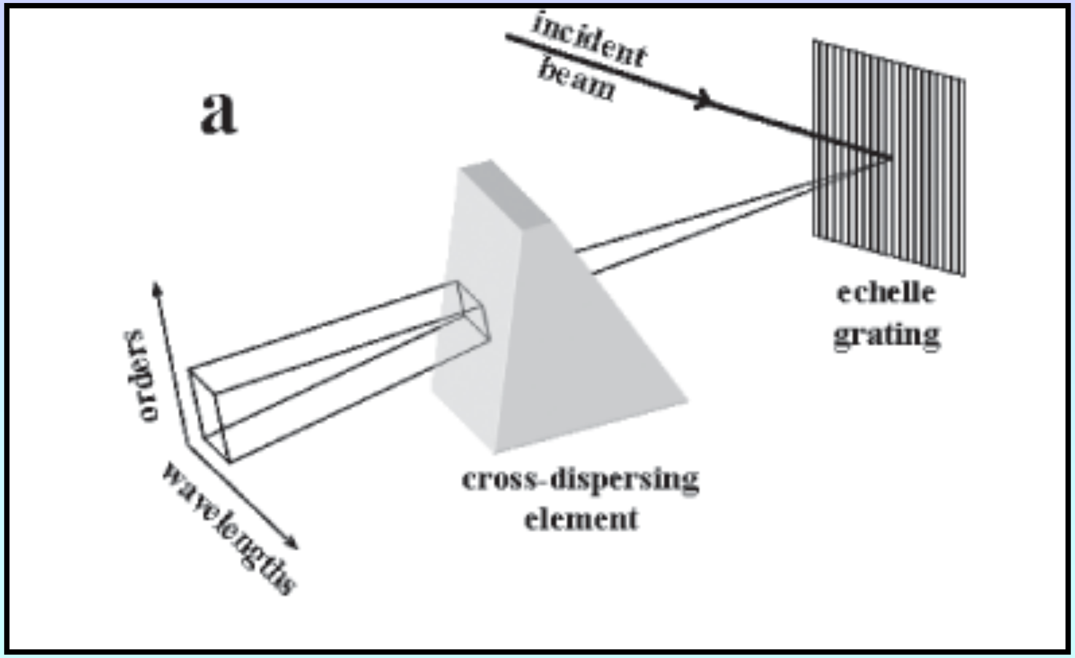


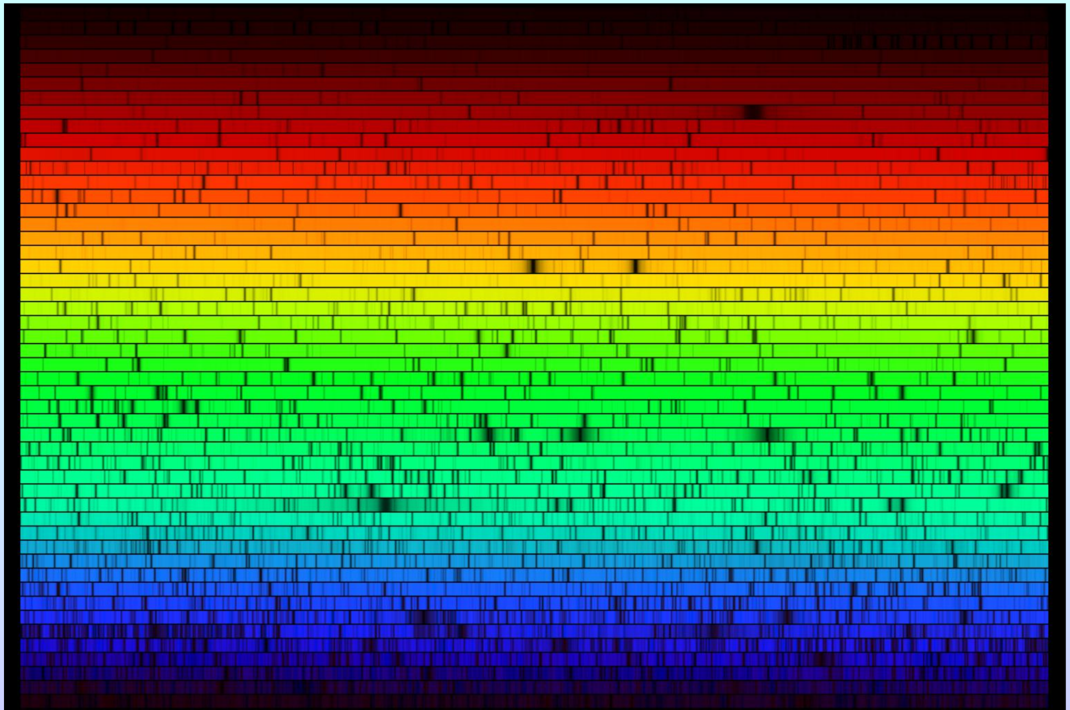


Echelle Spectrographs

Some gratings are optimized for high incidence angles, and therefore high diffraction orders. The output can then be cross-dispersed to produce a two-dimensional spectrum. This is called an echelle spectrograph.

Echelle spectrographs are typically used for high dispersion work.





Type	Advantages	Disadvantages
Narrow-band Imaging	Entire field observed at once. Total flux recorded.	Only one wavelength covered.
Objective Prism/Grism	Entire field observed at once. Total flux recorded.	High background; restricted to relatively low dispersion.
Long-slit Spectroscopy	Good estimate of local sky. Slit can be oriented at the parallactic angle.	Only one (or two) objects at a time. Sky relatively well measured. Slit losses. Possible issues of stability.
Multi-slit Spectroscopy	Can measure ~ 20 objects at once.	Choices limited by overlapping spectra; limited sky area; possible issues with atmospheric dispersion. Slit losses. Possible issues of stability.
Multi-fiber Spectroscopy	Can measure ~1000 objects at once. Stable spectrograph in controlled environment.	No estimate of local sky; affected by atmospheric dispersion. Slit and fiber losses. Throughput differences for each fiber.
Integral-Field Spectroscopy	Complete area coverage. Stable spectrograph in controlled environment.	Filling factor issues (unless lenslets attached to each fiber). Throughput differences for each fiber.
Echelle Spectroscopy	Very high dispersion with complete wavelength coverage.	One only object at a time. Slit losses. Usually for high resolution.

Very Basic Statistics

There are 3 types of statistical distributions that are frequently encountered in astronomy

- Binomial distributions
- Gaussian distributions
- Poisson distributions

Binomial Distribution

The binomial distribution applies to systems where there are two possible outcomes, with probability p for the first (positive) outcome and probability $1 - p$ for the alternative. The probability of observing n positive outcomes out of a sample of N experiments is

$$P(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

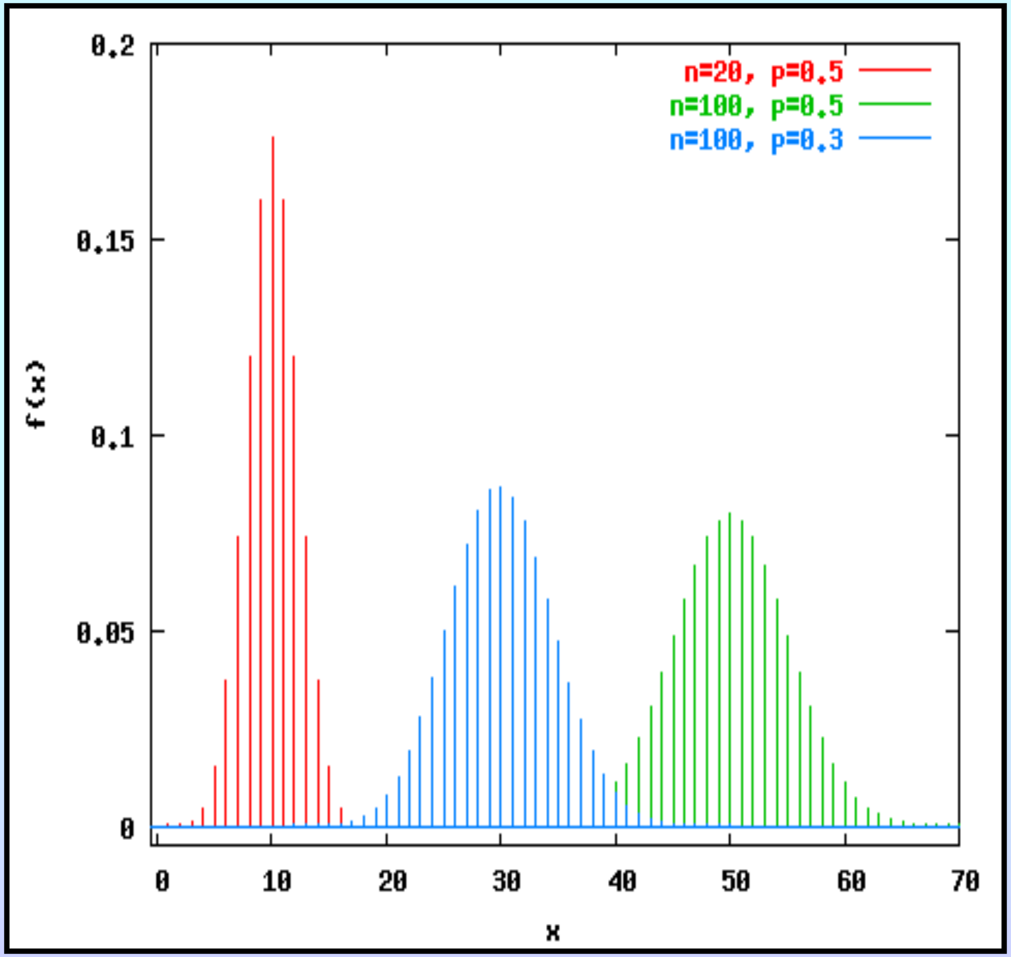
The mean of this distribution is

$$\langle n \rangle = Np$$

and the variance is

$$\sigma^2 = Np(1-p)$$

As with all probability distributions, this normalizes to one.

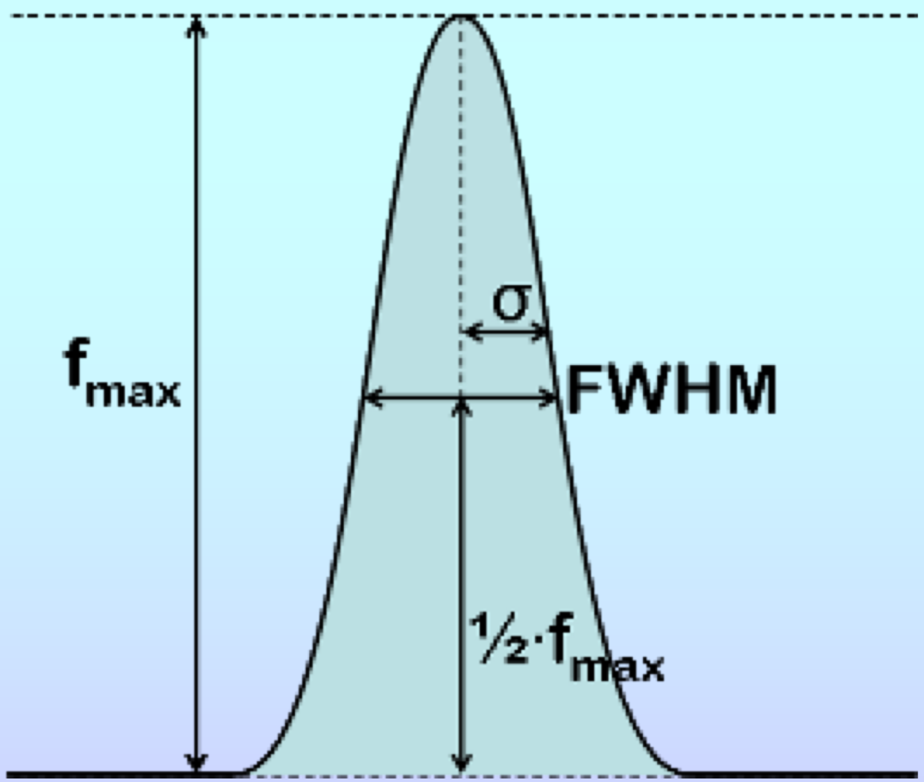


Gaussian Distribution

As $N \to \infty$, and Np much greater than 1, the binomial distribution converges to the Gaussian distribution,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

μ = mean
σ^2 = variance

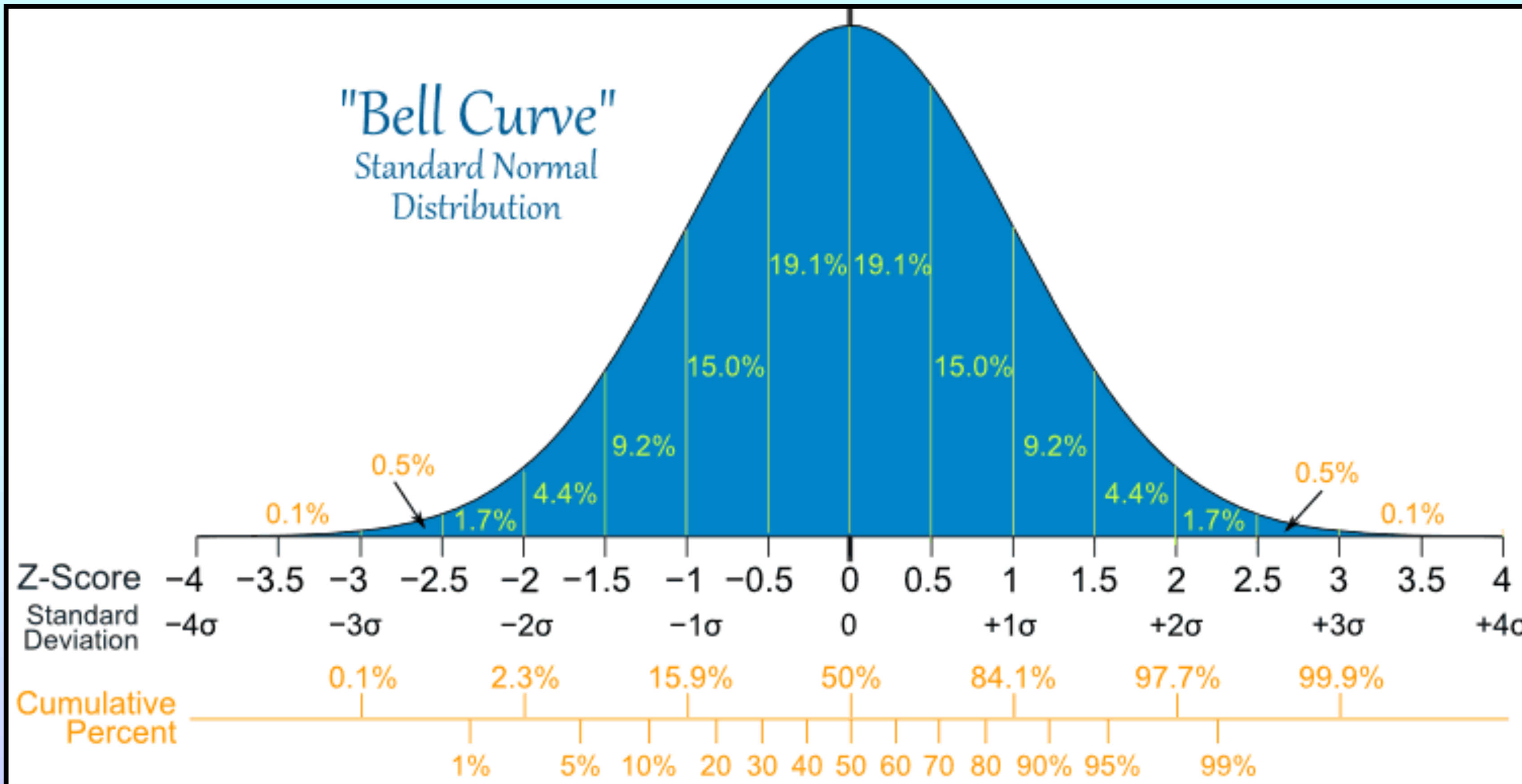


The variable σ, defined as the square root of the variance, is the "standard deviation". It is *not* the full-width-at-half-maximum (FWHM). The two are related by $\Gamma = 2.354\,\sigma$

Gaussian Distribution

The area under the curve of the Gaussian distribution determines the probability of obtaining a value within $\pm\, z$ of the mean. It is related to the "error function" by

$$\int_{-z}^{z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \operatorname{erf}\left(\frac{z}{\sigma\sqrt{2}}\right) \quad \text{where} \quad \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$$



Gaussian Distribution

The area under the curve of the Gaussian distribution determines the probability of obtaining a value within $\pm\, z$ of the mean. It is related to the "error function" by

$$\int_{-z}^{z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \operatorname{erf}\left(\frac{z}{\sigma\sqrt{2}}\right) \quad \text{where} \quad \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$$

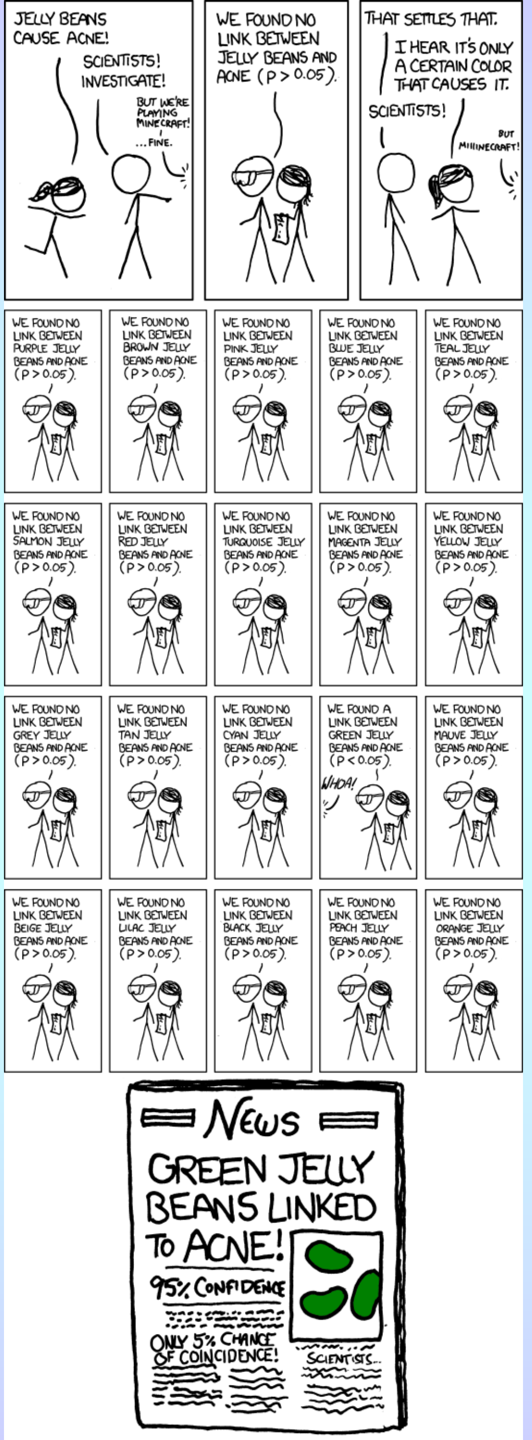
z	Prob	z	Prob
0.5	0.38292	3.0	0.9973002
1.0	0.68269	3.5	0.99953474
1.5	0.86638	4.0	0.999936656
2.0	0.95449	4.5	0.9999932043
2.5	0.98758	5.0	0.99999942657

Often, one also sees tables of erfc, which gives the complement of the error function.

Statistical Inference

Note: although some of these probabilities seem low, it all depends on the specifics of the experiment. For example, a *single* CCD may have ~16,000,000 pixels (and a Mosaic of CCDs will have many times more). In this case, 5 σ events will occur!

z	Prob	z	Prob
0.5	0.38292	3.0	0.9973002
1.0	0.68269	3.5	0.99953474
1.5	0.86638	4.0	0.999936656
2.0	0.95449	4.5	0.9999932043
2.5	0.98758	5.0	0.99999942657



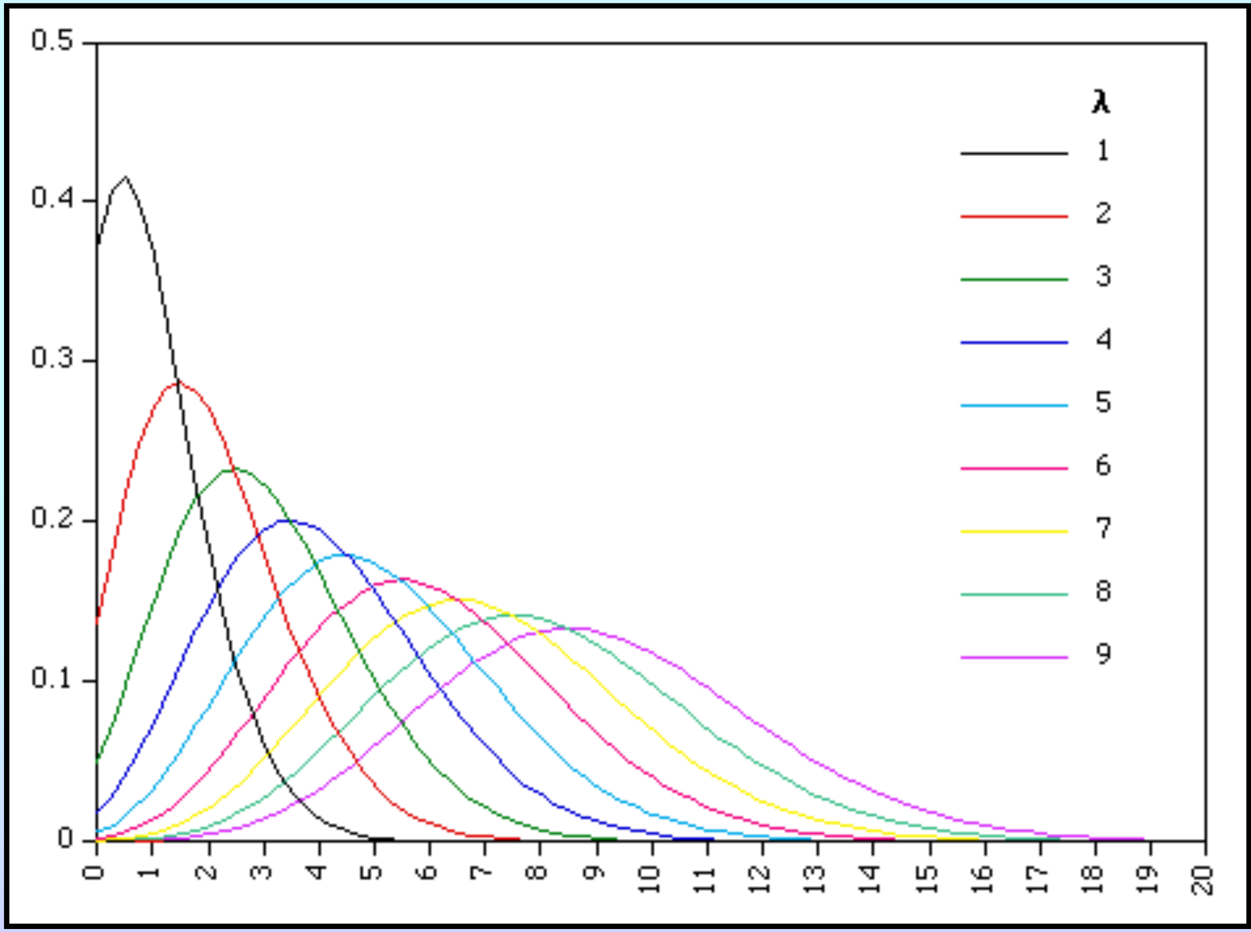
Poisson Distribution

As $N \rightarrow \infty$, and $p \rightarrow 0$, the binomial distribution approaches the Poisson distribution,

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad \lambda = \text{mean} = \text{variance}$$

As the mean increases, the Poisson distribution approaches the Gaussian distribution.

Poisson statistics are often called "counting statistics" or "shot noise," due to the discrete nature of the events (photon or electron counts).



Very Small Number Statistics

[Gehrels 1986, ApJ, 303, 336]

[Kraft et al. 1991, ApJ 374, 344]

Photons counts obey Poisson statistics. If N photons are received in one time interval, then in the next time interval, one might expect to detect $N \pm \sqrt{N}$ photons. But this is not quite true, since the first time interval also received $N \pm \sqrt{N}$ photons. (In other words, N is not known; we only measure $N \pm \sqrt{N}$.) Still, except when dealing with very few counts, one usually assumes $N_{\text{obs}} = N$.

Propagation of Errors

If a measurement, $x = f(u,v,w)$, is the combination of several variables, u, v, w, etc., each with its own uncertainty (σ_u, σ_v, σ_w, etc.), then the resulting uncertainty in x will be

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \sigma_w^2 \left(\frac{\partial x}{\partial w}\right)^2 - 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \ldots$$

The last set of terms represent the effect of covariance between the variables. If u, v, w, etc., are independent variables, then those terms are zero. So, for example, if $x = a \ln (b\, u)$, then

$$\frac{\partial x}{\partial u} = \frac{a}{u} \quad \Rightarrow \quad \sigma_x^2 = \frac{a^2}{u^2}\,\sigma_u^2$$

Standard Deviation and Standard Deviation of the Mean

Standard deviation, σ, describes how a set of measurements are distributed. If σ is known, then the measurement of a quantity X is generally quoted as X ± σ.

If multiple measurements are made, then the knowledge of X improves. Specifically, if *N* independent measurements are made, then the accuracy of X improves as $\sqrt{N}$.

Thus, there are two quantities that are often called "standard deviation": the true standard deviation, σ, and the standard deviation *of the mean*. One is related to the other by

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{N}}$$

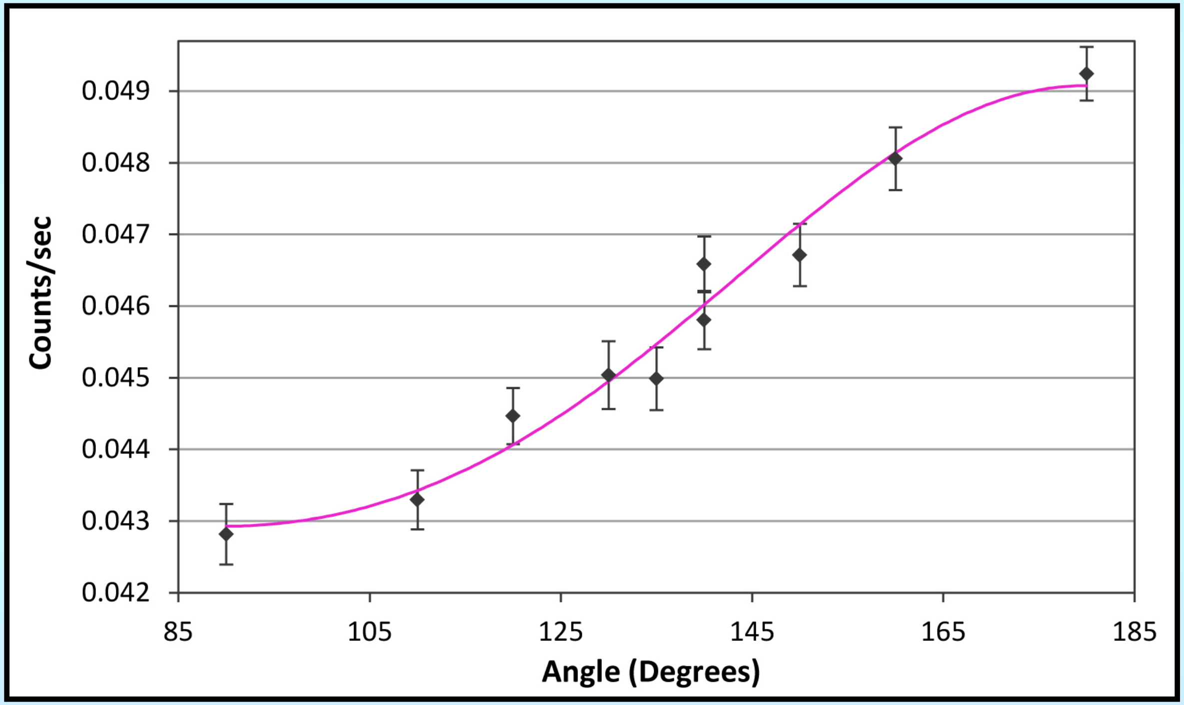
Be careful with your terminology! Many times, it is not clear which value the author is talking about!

Two Types of Statistics

Many times, one wants to fit a function to a set of data. There are two ways to do this: through a least-squares "frequentist" approach, or through a probabilistic "Bayesian" analysis. Often, a problem that is difficult for one technique is easy for the other.

Minimizing χ^2

The frequentist approach is characterized by minimizing a value such as χ^2; for a given set of data (x_i, y_i) with uncertainties σ_i, one compares the observed data to a model, $f(x)$, via

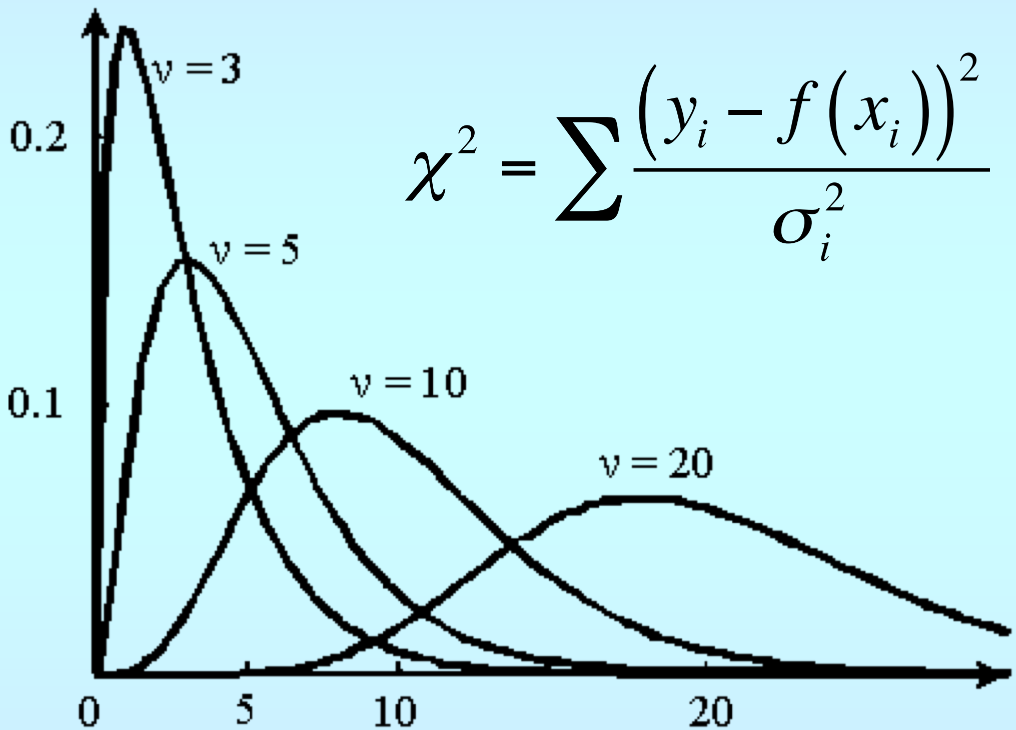


$$\chi^2 = \sum \frac{\left(y_i - f\left(x_i\right)\right)^2}{\sigma_i^2}$$

Of course, this approach is based on a bunch of assumptions (no errors in x, normal distributions, etc.)

Minimizing χ^2

If the errors are Gaussian, and entirely in x, the χ^2 statistic can be directly translated into a probability that the model is incorrect.



$$\chi^2 = \sum \frac{\left(y_i - f\left(x_i\right)\right)^2}{\sigma_i^2}$$

ν = degrees of freedom. This is usually the number of data points minus the number of parameters to be fit. Also often quoted are "reduced χ^2" values, which are χ^2/ν.

(Even if the errors are not Gaussian, the best fit is (most times) that which minimizes χ^2. However, in this case, χ^2 cannot be translated into a probability.)

Bayesian Methods

Assume that a set of data can be fit with a function $f(x)$. In the interval between x_i and $x_i + \Delta x$, one might expect to find $\lambda = f(x)\,\Delta x$ objects. However, if Poisson statistics hold, the probability of actually observing n objects in the interval is

$$p_i(n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

The same is true for all the other intervals, so the likelihood of observing n_1 objects in bin 1, n_2 in bin 2, etc., is $L = \prod_i p_i(n_i)$

Now choose a different set of parameters (or different function entirely) and repeat the process. You can then state how much more (or less) likely the first function is than the second. The process of computing (relative) likelihoods can also be extended by adding "priors" derived from other data sets, etc.